

Framework Well-Architected da AWS

Pilar da otimização de custos



Pilar da otimização de custos: Framework Well-Architected da AWS

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens comerciais da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestige a Amazon. Todas as outras marcas comerciais que não pertencem à Amazon pertencem a seus respectivos proprietários, que podem ou não ser afiliados, patrocinados pela Amazon ou ter conexão com ela.

Table of Contents

Resumo e introdução	1
Resumo	1
Introdução	1
Otimização de custo	3
Princípios de design	3
Definição	4
Gerenciamento financeiro na nuvem	5
COST01-BP01 Estabelecer a propriedade da otimização de custos	8
Orientação para implementação	8
Recursos	10
COST01-BP02 Estabelecer uma parceria entre finanças e tecnologia	11
Orientação para implementação	8
Recursos	10
COST01-BP03 Estabelecer orçamentos e previsões para a nuvem	16
Orientação para implementação	8
Recursos	10
COST01-BP04 Implementar a conscientização de custos em seus processos organizacionais	20
Orientação para implementação	8
Recursos	10
COST01-BP05 Relatar e notificar sobre a otimização de custos	22
Orientação para implementação	8
Recursos	10
COST01-BP06 Monitorar custos proativamente	25
Orientação para implementação	8
Recursos	10
COST01-BP07 Manter-se em dia com os novos lançamentos de serviços	27
Orientação para implementação	8
Recursos	10
COST01-BP08 Criar uma cultura de conscientização de custos	28
Orientação para implementação	8
Recursos	10
COST01-BP09 Quantificar o valor comercial proveniente da otimização de custos	31
Orientação para implementação	8

Recursos	10
Reconhecimento de despesas e usos	34
Governança	34
COST02-BP01 Desenvolver políticas com base nos requisitos da sua organização	35
COST02-BP02 Implementar metas e objetivos	39
COST02-BP03 Implementar uma estrutura de contas	45
COST02-BP04 Implementar grupos e perfis	49
COST02-BP05 Implementar controles de custos	51
COST02-BP06 rastrear o ciclo de vida do projeto	54
Monitorar custos e uso	57
COST03-BP01 Configurar fontes de informações detalhadas	57
COST03-BP02 Adicionar informações da organização aos custos e ao uso	60
COST03-BP03 Identificar categorias de atribuição de custos	62
COST03-BP04 Estabelecer métricas da organização	66
COST03-BP05 Configurar ferramentas de faturamento e gerenciamento de custos	67
COST03-BP06 Alocar custos com base nas métricas de workload	70
Desativar recursos	72
COST04-BP01 Rastrear os recursos ao longo da vida útil	72
COST04-BP02 Implementar um processo de desativação	74
COST04-BP03 Desativar recursos	77
COST04-BP04 Desativar recursos automaticamente	77
COST04-BP05 Impor políticas de retenção de dados	79
Recursos economicamente eficientes	81
Avaliar o custo ao selecionar serviços	81
COST05-BP01 Identificar os requisitos de custos da organização	81
COST05-BP02 Analisar todos os componentes da workload	83
COST05-BP03 Executar uma análise completa de cada componente	86
COST05-BP04 Selecionar software com licenciamento econômico	88
COST05-BP05 Selecionar os componentes desta workload para otimizar o custo alinhado com as prioridades da organização	90
COST05-BP06 Realizar análises de custos para diferentes usos ao longo do tempo	92
Selecionar o tipo de recurso, o tamanho e o número corretos	94
COST06-BP01 Realizar modelagem de custos	95
COST06-BP02 Selecionar o tipo, o tamanho e o número do recurso com base nos dados	97
COST06-BP03 Selecionar o tipo, tamanho e número do recurso automaticamente com base nas métricas	99

COST06-BP04 Considerar o uso de recursos compartilhados	102
Selecionar o melhor modelo de preços	105
COST07-BP01 Executar análise de modelo de preço	111
COST07-BP02 Escolher regiões com base no custo	114
COST07-BP03 Selecionar contratos de terceiros com termos econômicos	116
COST07-BP04 Implementar modelos de preços para todos os componentes da workload ..	118
COST07-BP05 Realizar análise de modelo de preços em nível da conta de gerenciamento	120
Planejar a transferência de dados	123
COST08-BP01 Executar a modelagem de transferência de dados	123
COST08-BP02 Selecionar os componentes para otimizar o custo da transferência de dados	126
COST08-BP03 Implementar serviços para reduzir custos de transferência de dados	128
Gerenciar recursos de demanda e fornecimento	131
COST09-BP01 Realizar uma análise sobre a demanda da workload	131
Orientação para implementação	8
Recursos	10
COST09-BP02 Implementar um buffer ou controle de utilização para gerenciar a demanda	134
Orientação para implementação	8
Recursos	10
COST09-BP03 Fornecer recursos dinamicamente	137
Orientação para implementação	8
Etapas de implementação	9
Recursos	10
Otimização ao longo do tempo	145
Definir um processo de revisão e analisar sua workload com frequência	145
COST10-BP01 Desenvolver um processo de revisão de workloads	145
COST10-BP02 Revisar e analisar a workload regularmente	148
Automatização de operações	149
COST11-BP01 Realizar automações nas operações	150
Conclusão	155
Colaboradores	156
Outras fontes de leitura	157
Revisões do documento	158
Avisos	160
Glossário da AWS	161

Pilar Otimização de custos: AWS Well-Architected Framework

Data de publicação: 27 de junho de 2024 ([Revisões do documento](#))

Resumo

Este whitepaper destaca o pilar Otimização de custos do Amazon Web Services (AWS) Well-Architected Framework. Ele fornece orientações para ajudar os clientes a aplicar as práticas recomendadas nas áreas de projeto, entrega e manutenção de ambientes da AWS.

Uma workload com custo otimizado utiliza integralmente todos os recursos, alcança um resultado com o menor preço possível e atende a seus requisitos funcionais. Este whitepaper fornece orientações detalhadas para a criação de recursos dentro da organização, o projeto da workload, a seleção dos serviços, a configuração e a operação dos serviços, além da aplicação de técnicas de otimização de custos.

Introdução

O [AWS Well-Architected Framework](#) ajuda a entender as decisões tomadas no momento de criar workloads na AWS. O Framework fornece as práticas recomendadas de arquitetura para projetar e operar workloads confiáveis, seguras, eficientes, econômicas e sustentáveis na nuvem. Ele demonstra uma maneira de avaliar consistentemente suas arquiteturas em relação às práticas recomendadas e identificar áreas de melhoria. Acreditamos que ter as workloads bem arquitetadas aumenta muito a probabilidade de sucesso nos negócios.

O framework é baseado em seis pilares:

- Excelência operacional
- Segurança
- Confiabilidade
- Eficiência de performance
- Otimização de custo
- Sustentabilidade

Este documento concentra-se no pilar Otimização de custos e em como projetar workloads com o uso mais eficaz de serviços e recursos para atingir resultados de negócio com o menor preço.

Você aprenderá a aplicar as práticas recomendadas do pilar Otimização de custos em sua organização. A otimização de custos pode ser desafiadora em soluções on-premises, pois é necessário prever a capacidade futura e as necessidades de negócio enquanto realiza processos de aquisição complexos. A adoção das práticas neste documento ajudará sua organização a atingir estes objetivos:

- Gerenciamento financeiro na nuvem
- Reconhecimento de despesas e usos
- Recursos economicamente eficientes
- Gerenciar recursos de demanda e fornecimento
- Otimização ao longo do tempo

Este documento destina-se a profissionais de funções de tecnologia e finanças, como diretores de tecnologia (CTOs), diretores financeiros (CFOs), arquitetos, desenvolvedores, controladores financeiros, planejadores financeiros, analistas de negócios e membros da equipe de operações. Este documento não fornece detalhes de implementação ou padrões de arquitetura. No entanto, inclui referências a recursos apropriados.

Otimização de custo

A otimização de custos é um processo contínuo de refinamento e melhoria durante o período do ciclo de vida de uma workload. As práticas deste documento ajudam você a criar e operar workloads com reconhecimento de custo que alcançam resultados de negócios, minimizando custos e permitindo que sua organização maximize o retorno sobre o investimento.

Tópicos

- [Princípios de design](#)
- [Definição](#)

Princípios de design

Considere os seguintes princípios de projeto para otimização de custos:

Implementar o gerenciamento financeiro na nuvem: para obter sucesso financeiro e acelerar a realização de valor empresarial na nuvem, você deve investir em gerenciamento financeiro na nuvem. Sua organização deve dedicar o tempo e os recursos necessários para criar aptidão nesse novo domínio de gerenciamento de utilização e tecnologia. Semelhante à sua capacidade de Segurança ou Operações, você precisa criar capacidade por meio da criação de conhecimento, programas, recursos e processos para ajudá-lo a se tornar uma organização econômica.

Adotar um modelo de consumo: pague apenas pelos recursos de computação que você consome e aumente ou diminua o uso dependendo dos requisitos da empresa. Por exemplo, ambientes de desenvolvimento e teste são geralmente usados apenas por oito horas ao dia durante a semana de trabalho. É possível parar esses recursos quando eles não estiverem em uso para obter uma redução de custos potencial de 75% (40 horas versus 168 horas).

Medir a eficiência geral: meça o resultado de negócios da workload e os custos associados à sua entrega. Use esses dados para entender os ganhos obtidos com o aumento da saída, o aumento da funcionalidade e a redução de custos.

Parar de gastar dinheiro em tarefas pesadas genéricas: a AWS faz o trabalho pesado das operações de datacenter, como o armazenamento em rack, o empilhamento e a alimentação de servidores. Ela também elimina a sobrecarga operacional do gerenciamento de sistemas operacionais e aplicações com serviços gerenciados. Isso permite que você mantenha o foco em seus clientes e projetos de negócios, e não na infraestrutura de TI.

Analisar e atribuir gastos: a nuvem facilita a identificação precisa do custo e uso das workloads, o que permite a atribuição transparente de custos de TI para fluxos de receita e proprietários de workloads individuais. Dessa forma, a medição do retorno sobre o investimento (ROI) é facilitada e os proprietários de workloads têm a oportunidade de otimizar recursos e reduzir custos.

Definição

Há cinco áreas de foco para otimização de custos na nuvem:

- Gerenciamento financeiro na nuvem
- Reconhecimento de despesas e usos
- Recursos economicamente eficientes
- Gerenciamento de recursos de demanda e oferta
- Otimização ao longo do tempo

De forma semelhante aos outros pilares do Well-Architected Framework, há comprometerimentos que devem ser levados em conta para a otimização de custos. Por exemplo, otimizar para velocidade de entrada no mercado ou para custos. Em alguns casos, é melhor otimizar a velocidade para entrar no mercado rapidamente, enviar novos recursos ou cumprir um prazo, em vez de investir na otimização de custos inicial.

Às vezes, as decisões de projeto são tomadas com base na pressa e não em dados, já que sempre existe a tentação de compensar excessivamente, em vez de dedicar tempo a realizar benchmarking da implantação mais econômica. A compensação excessiva pode levar a implantações com provisionamento excessivo e não otimizado. No entanto, pode ser uma escolha razoável se você precisa mover sem alterações (lift-and-shift) recursos de seu ambiente on-premises para a nuvem e então realizar a otimização posteriormente.

Investir na quantidade certa de esforço em uma estratégia de otimização de custos com antecedência permite aproveitar os benefícios econômicos da nuvem de modo mais rápido, garantindo uma adesão consistente às práticas recomendadas e evitando provisionamento excessivo desnecessário. As seções a seguir fornecem técnicas e práticas recomendadas para a implementação inicial e contínua do gerenciamento financeiro na nuvem e otimização de custos para suas workloads.

Gerenciamento financeiro na nuvem

O gerenciamento de finanças na nuvem requer a evolução de seus processos financeiros existentes para estabelecer e operar com transparência de custos, controle, planejamento e otimização para seus ambientes do AWS.

A aplicação de modelos tradicionais e estáticos de planejamento em cascata, orçamento de TI e avaliação de custos ao uso dinâmico da nuvem pode criar riscos, levar a um planejamento impreciso e resultar em menos visibilidade. Em última análise, isso resulta em uma oportunidade perdida de otimizar e controlar os custos com eficiência e obter valor comercial de longo prazo. Para evitar esses obstáculos, gerencie ativamente os custos em toda a jornada para a nuvem, esteja você criando aplicações nativamente na nuvem, migrando suas workloads para a nuvem ou expandindo sua adoção de serviços em nuvem.

O gerenciamento financeiro na nuvem (CFM) permite que organizações financeiras, de produtos, de tecnologia e de negócios gerenciem, otimizem e planejem custos à medida que aumentam seu uso e mudam seu dimensionamento na AWS. O principal objetivo do CFM é permitir que os clientes alcancem seus resultados empresariais da maneira mais econômica e acelerem a criação de valor econômico e de negócios, encontrando o equilíbrio certo entre agilidade e controle.

As soluções de CFM ajudam a transformar seus negócios por meio de transparência, controle, previsão e otimização de custos. Essas soluções também podem criar uma cultura consciente de custos que impulsiona a responsabilidade em todas as equipes e funções. As equipes financeiras podem ver de onde vêm os custos, executar operações com o mínimo de despesas inesperadas, planejar o uso dinâmico da nuvem e economizar nas despesas da nuvem enquanto as equipes dimensionam suas adoções na nuvem. O compartilhamento disso com as equipes de engenharia pode fornecer o contexto financeiro necessário para a seleção, o uso e a otimização de recursos.

O CFM da AWS oferece um conjunto de recursos para gerenciar, otimizar e planejar os custos da nuvem, mantendo a agilidade dos negócios. O CFM é fundamental não apenas para gerenciar os custos com eficácia, mas também para verificar se os investimentos estão gerando os resultados comerciais esperados. Esses são os quatro pilares do framework de gerenciamento financeiro na nuvem adotados pela Nuvem AWS: ver, economizar, planejar e executar. Cada um desses pilares possui um conjunto de atividades e capacidades.



Os quatro pilares do gerenciamento financeiro na nuvem.

- Ver: como você está medindo, monitorando e gerando responsabilidade por seus gastos com a nuvem no momento? Se você é novo na AWS ou planeja usar a AWS, tem algum plano para estabelecer visibilidade de custos e uso?

Para entender seus custos na AWS e otimizar os gastos, você precisa saber a origem desses custos. Isso requer uma estrutura deliberada para suas contas e recursos, ajudando sua organização financeira a rastrear os fluxos de gastos e responsabilizar as equipes por suas partes nos resultados.

Serviços da AWS: AWS Control Tower, AWS Organizations, tags de alocação de custos, políticas de tags, AWS Resource Groups, Categorias de Custos da AWS, AWS Cost Explorer, AWS Cost and Usage Report, IRs e SPs

Recursos: Práticas recomendadas de marcação com tags da AWS, Categorias de Custos da AWS

- Economizar: quais alavancas de otimização de custos você está usando para otimizar seus gastos? Se você não usa a AWS, está familiarizado com as otimizações comuns baseadas em modelos de preços e uso?

No princípio da economia, otimizamos os custos com recomendações de recursos e preços. A otimização de custos começa com uma estratégia bem definida para seu novo modelo operacional na nuvem. O ideal é que isso comece o mais cedo possível em sua jornada para a nuvem, preparando o terreno para uma cultura consciente de custos reforçada pelos processos e comportamentos corretos.

Há muitas maneiras diferentes de otimizar os custos da nuvem. Uma delas é selecionar o modelo de compra certo (IRs e SPs) ou se sua workload é imutável e está em contêiner para que você possa adotar instâncias spot do Amazon EC2. Além disso, escale sua workload usando grupos do Amazon EC2 Auto Scaling.

Serviços da AWS: IRs e SPs, grupos do Amazon EC2 Auto Scaling, instâncias spot

Recursos: Instâncias reservadas, Savings Plans, práticas recomendadas para lidar com o Amazon EC2

- Planejar: como você planeja atualmente o uso e os gastos futuros na nuvem? Você tem uma metodologia para quantificar a geração de valor para uma nova migração? Você atualizou seus processos atuais de orçamento e previsão para adotar o uso variável da nuvem?

O princípio do planejamento significa melhorar seu planejamento com orçamento e previsão flexíveis. Depois de estabelecer a visibilidade e os controles de custo, você provavelmente desejará planejar e definir as expectativas de gastos em projetos de nuvem. O AWS oferece flexibilidade para criar processos dinâmicos de previsão e orçamento para que você possa saber se os custos atendem ou excedem os limites orçamentários.

Serviços da AWS: AWS Cost Explorer, AWS Cost and Usage Report, AWS Budgets

Recursos: previsões baseadas no uso, relatórios do AWS Budgets e alertas

- Executar: quais são alguns dos processos e ferramentas operacionais que você está usando para gerenciar seus gastos com a nuvem e quem está liderando esses esforços? Você já pensou em como as coisas funcionarão do ponto de vista das operações diárias assim que começar a usar a AWS?

O princípio da execução é, na verdade, gerenciar o faturamento e o controle de custos. É possível estabelecer barreiras de proteção e definir a governança para garantir que as despesas fiquem alinhadas com os orçamentos. A AWS fornece várias ferramentas para ajudar você a começar.

Serviços da AWS: console do AWS Billing and Cost Management, AWS Identity and Access Management, políticas de controle de serviços (SCP), AWS Service Catalog, AWS Cost Anomaly Detection, AWS Budgets

Recursos: Conceitos básicos do console do AWS Billing

As seguintes práticas são recomendadas para o gerenciamento financeiro na nuvem:

Práticas recomendadas

- [COST01-BP01 Estabelecer a propriedade da otimização de custos](#)
- [COST01-BP02 Estabelecer uma parceria entre finanças e tecnologia](#)
- [COST01-BP03 Estabelecer orçamentos e previsões para a nuvem](#)
- [COST01-BP04 Implementar a conscientização de custos em seus processos organizacionais](#)
- [COST01-BP05 Relatar e notificar sobre a otimização de custos](#)
- [COST01-BP06 Monitorar custos proativamente](#)
- [COST01-BP07 Manter-se em dia com os novos lançamentos de serviços](#)
- [COST01-BP08 Criar uma cultura de conscientização de custos](#)
- [COST01-BP09 Quantificar o valor comercial proveniente da otimização de custos](#)

COST01-BP01 Estabelecer a propriedade da otimização de custos

Crie uma equipe (escritório de negócios na nuvem, Centro de Excelência da Nuvem ou FinOps) responsável por estabelecer e manter a conscientização de custos em toda a organização. O responsável pela otimização de custos pode ser uma pessoa ou uma equipe (requer pessoal das equipes de finanças, tecnologia e negócios) que conheça toda a organização e as finanças da nuvem.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Esta é a introdução de uma função ou equipe de escritório de negócios na nuvem (CBO) ou Centro de Excelência da Nuvem (CCOE) responsável por estabelecer e manter uma cultura de conscientização de custos de computação em nuvem. Em toda a organização, essa função pode ser exercida por qualquer pessoa ou equipe existente, ou por uma nova equipe com as principais partes interessadas em finanças, tecnologia e organização.

A função (individual ou equipe) prioriza e dedica a porcentagem necessária de seu tempo a atividades de gerenciamento e otimização de custos. Para uma organização pequena, a função pode gastar uma porcentagem de tempo menor em comparação com uma função de tempo integral para uma empresa maior.

A função exige uma abordagem multidisciplinar, com recursos de gerenciamento de projetos, ciência de dados, análise financeira e desenvolvimento de software ou infraestrutura. Ela pode melhorar a eficiência da workload realizando otimizações de custos em três propriedades diferentes:

- Centralizada: por meio de equipes designadas, como a equipe FinOps, a equipe de gerenciamento financeiro na nuvem (CFM), o escritório de negócios na nuvem (CBO) ou o Centro de Excelência da Nuvem (CCoE), os clientes podem projetar e implementar mecanismos de governança e promover as práticas recomendadas em toda a empresa.
- Descentralizada: as equipes de tecnologia são convencidas a realizar otimizações de custos.
- Híbrida: combinação de equipes centralizadas e descentralizadas que podem trabalhar em conjunto para realizar otimizações de custo.

A função pode ser medida ao comparar a sua capacidade de realização e entrega com as metas de otimização de custos (por exemplo, métricas de eficiência da workload).

É necessário garantir que haja patrocínio executivo para essa função, o que é um fator de sucesso fundamental. O patrocinador é considerado defensor do consumo de nuvem econômico e oferece suporte ao encaminhamento para a equipe a fim de garantir que as atividades de otimização de custos sejam tratadas de acordo com o nível de prioridade definido pela organização. Caso contrário, a orientação poderá ser ignorada e as oportunidades de redução de custo não serão priorizadas. Juntos, o patrocinador e a equipe ajudam a organização a consumir a nuvem com eficiência e agregar valor comercial.

Se você tem um [plano de suporte](#) Business, Enterprise-On-Ramp ou Enterprise Support e precisa de ajuda para elaborar essa equipe ou função, entre em contato com seus especialistas de gerenciamento financeiro na nuvem (CFM) por meio de sua equipe de conta.

Etapas de implementação

- Defina os membros principais: todas as partes relevantes da organização devem contribuir e ter interesse pelo gerenciamento de custos. As equipes comuns dentro das organizações geralmente incluem: finanças, proprietários de aplicações ou produtos, gerenciamento e equipes técnicas (DevOps). Alguns são contratados em tempo integral (financeiro ou técnico), enquanto outros são contratados periodicamente, conforme necessário. Pessoas ou equipes encarregadas de executar o CFM precisam dos seguintes conjuntos de habilidades:
 - Desenvolvimento de software: quando ocorre o desenvolvimento de scripts e automação.

- Engenharia de infraestrutura: para implantar scripts, automatizar processos e entender como os serviços e os recursos são provisionados.
- Perspicácia de operações: CFM é sobre operar na nuvem de maneira eficiente por meio de medição, monitoramento, modificação, planejamento e dimensionamento do uso eficiente da nuvem.
- Defina objetivos e métricas: a função precisa agregar valor à organização de diferentes formas. Esses objetivos são definidos e evoluem continuamente com a organização. As atividades comuns incluem: criação e execução de programas educacionais sobre otimização de custos em toda a organização, desenvolvimento de padrões em toda a organização (como monitoramento e geração de relatórios para otimização de custos) e definição de metas de workload sobre otimização. Essa função também precisa informar regularmente a organização sobre o recurso de otimização de custos.

Você pode definir indicadores-chave de performance (KPIs) baseados em valor ou custo. Ao definir os KPIs, você pode calcular o custo esperado em termos de eficiência e o resultado comercial esperado. KPIs baseados em valor vinculam métricas de uso e custo a motivadores de valor empresarial e ajudam a racionalizar mudanças em gastos na AWS. O primeiro passo para derivar KPIs baseados em valor é trabalhar em conjunto, em toda a organização, para selecionar e concordar sobre um conjunto padrão de KPIs.

- Estabeleça uma cadência regular: o grupo (equipes financeira, empresarial e de tecnologia) devem se reunir regularmente para analisar metas e métricas. Um ritmo típico envolve analisar o estado da organização, todos os programas em execução no momento e as métricas financeiras e de otimização gerais. Em seguida, as principais workloads são relatadas em mais detalhes.

Durante essas revisões regulares, é possível analisar a eficiência (custo) da workload e o resultado empresarial. Por exemplo, um aumento de 20% no custo de uma workload pode ser consequência de um aumento do uso pelos clientes. Neste caso, esse aumento de 20% no custo pode ser interpretado como um investimento. Essas chamadas regulares podem ajudar as equipes a identificar KPIs de valor que ofereçam propósito para toda a organização.

Recursos

Documentos relacionados:

- [Blog de CCoE da AWS](#)
- [Criar um escritório de negócios na nuvem](#)

- [CCoE: Centro de Excelência da Nuvem](#)

Vídeos relacionados:

- [História de sucesso de CCoE de vanguarda](#)

Exemplos relacionados:

- [Usar um Centro de Excelência da Nuvem \(CCoE\) para transformar toda a empresa](#)
- [Criar um CCoE para transformar toda a empresa](#)
- [Sete obstáculos que devem ser evitados ao criar um CCoE](#)

COST01-BP02 Estabelecer uma parceria entre finanças e tecnologia

Envolva equipes financeiras e de tecnologia em discussões sobre custo e uso em todas as etapas da jornada para a nuvem. As equipes se reúnem e discutem regularmente assuntos como objetivos e metas organizacionais, o estado atual de custo e uso e práticas financeiras e contábeis.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

As equipes de tecnologia inovam mais rapidamente na nuvem devido à redução dos ciclos de implantação de aprovação, aquisição e infraestrutura. Isso pode ser um ajuste para organizações financeiras anteriormente usadas para executar processos demorados e com uso intensivo de recursos para aquisição e implantação de capital em ambientes de datacenter on-premises, além de alocação de custos apenas na aprovação do projeto.

Do ponto de vista da organização financeira e de aquisição, o processo de definição orçamentária, solicitações de capital, aprovações, aquisição e instalação de infraestrutura física é algo que levou décadas para ser aprendido e padronizado:

- Equipes de engenharia ou TI costumam ser os solicitantes
- Equipes financeiras diversas atuam como aprovadores e compradores
- Equipes de operação estendem, acumulam e disponibilizam infraestrutura pronta para ser usada



Com a adoção da nuvem, a aquisição e o consumo de infraestrutura deixaram de estar vinculados a uma série de dependências. No modelo de nuvem, as equipes de tecnologia e produto deixam de ser simples desenvolvedoras, passando a ser operadoras e proprietárias de seus produtos, além de responsáveis pela maioria das atividades historicamente associadas às equipes financeiras e de operações, incluindo aquisição e implantação.

Basta uma conta e o conjunto adequado de permissões para provisionar recursos na nuvem. Também é isso que reduz o risco financeiro e de TI, o que significa que as equipes estão sempre a poucos cliques ou chamadas de API de encerrar recursos ociosos ou desnecessários na nuvem. Também é isso que permite que as equipes de tecnologia inovem com mais rapidez: a agilidade e capacidade de aplicar e derrubar experimentos. Embora a natureza variável do consumo na nuvem

possa afetar a previsibilidade do ponto de vista de previsão e definição orçamentária, a nuvem oferece às organizações a capacidade de reduzir o custo de provisionamento em excesso, além de reduzir o custo de oportunidade associado ao subprovisionamento conservador.



Estabelecer uma parceria entre as principais partes interessadas em finanças e tecnologia para criar uma compreensão compartilhada dos objetivos organizacionais e desenvolver mecanismos para obter sucesso financeiro no modelo de gastos variáveis da computação em nuvem. As equipes relevantes da sua organização devem estar envolvidas em discussões de custo e uso em todas as fases da jornada para a nuvem, incluindo:

- Líderes financeiros: CFOs, controladores financeiros, planejadores financeiros, analistas de negócios, aquisições, sourcing e contas a pagar devem compreender o modelo de nuvem de consumo, as opções de compra e o processo de faturamento mensal. O departamento financeiro precisa se unir às equipes de tecnologia para criar e socializar uma narrativa de valor de TI, ajudando as equipes comerciais a entender como o gasto com tecnologia está associado aos

resultados comerciais. Assim, as despesas com tecnologia são vistas não como custos, mas como investimentos. Devido às diferenças fundamentais entre a nuvem (como a taxa de alteração no uso, definição de preço com pagamento conforme o uso, definição de preço em camadas, modelos de definição de preço e informações detalhadas de faturamento e uso) em comparação à operação on-premises, é essencial que a organização financeira entenda como o uso da nuvem pode afetar aspectos empresariais, incluindo processos de aquisição, rastreamento de incentivos, alocação de custos e demonstrações financeiras.

- Líderes de tecnologia: os líderes de tecnologia (incluindo proprietários de produtos e aplicações) devem estar cientes dos requisitos financeiros (por exemplo, restrições orçamentárias), bem como dos requisitos de negócios (por exemplo, contratos de nível de serviço). Isso permite que a workload seja implementado para atingir os objetivos desejados da organização.

A parceria entre finanças e tecnologia oferece os seguintes benefícios:

- As equipes de finanças e tecnologia têm visibilidade praticamente em tempo real dos custos e do uso.
- As equipes de finanças e tecnologia estabelecem um procedimento operacional padrão para lidar com a variação de gastos na nuvem.
- As partes interessadas em finanças atuam como consultores estratégicos com relação à forma como o capital é usado para comprar descontos de compromissos (por exemplo, instâncias reservadas ou Savings Plans da AWS) e como a nuvem é usada para expandir a organização.
- Contas a pagar e processos de aquisição existentes são usados com a nuvem.
- As equipes de finanças e tecnologia colaboram na previsão de custos e uso futuros da AWS para alinhar e criar orçamentos organizacionais.
- Melhor comunicação entre organizações por meio de uma linguagem compartilhada e entendimento comum dos conceitos financeiros.

As partes interessadas adicionais dentro da sua organização que devem ser envolvidas em discussões de custo e uso incluem:

- Proprietários de unidades de negócios: os proprietários de unidades de negócios devem compreender o modelo de negócios de nuvem para que possam fornecer orientações tanto para as unidades de negócios quanto para toda a empresa. Esse conhecimento de nuvem é essencial quando há necessidade de prever o crescimento e o uso da workload, e ao avaliar opções de compra de longo prazo, como instâncias reservadas ou Savings Plans.

- Equipe de engenharia: uma parceria entre as equipes financeira e de tecnologia é essencial para o desenvolvimento de uma cultura de consciência dos custos que encoraja os engenheiros a agirem em relação ao gerenciamento financeiro na nuvem (CFM). Um dos problemas comuns dos profissionais de CFM ou operações financeiras e das equipes financeiras é fazer com que os engenheiros entendam todos os negócios na nuvem, sigam as práticas recomendadas e adotem as medidas recomendadas.
- Terceiros: se sua organização usa terceiros (por exemplo, consultores ou ferramentas), certifique-se de que eles estejam alinhados aos seus objetivos financeiros e possam demonstrar o alinhamento por meio de seus modelos de engajamento e um retorno sobre o investimento (ROI). Terceiros normalmente contribuirão para o relatório e a análise de qualquer workload que gerenciem e fornecerão análise de custo de qualquer workload que projetem.

Implementar o CFM e obter sucesso requer a colaboração das equipes financeira, comercial e de tecnologia, além de uma mudança na forma como os gastos com nuvem são comunicados e avaliados em toda a organização. Inclua as equipes de engenharia para que façam parte dessas conversas sobre custos e uso em todos os estágios, incentivando-as a seguir as práticas recomendadas e tomar medidas previamente acordadas conforme for apropriado.

Etapas de implementação

- Defina os membros importantes: verifique se todos os membros relevantes de suas equipes de finanças e tecnologia participam da parceria. Os membros financeiros relevantes serão aqueles que interagem com a conta da nuvem. Normalmente serão CFOs, controladores financeiros, planejadores financeiros, analistas de negócios, compras e sourcing. Normalmente, os membros de tecnologia serão proprietários de produtos e aplicações, gerentes técnicos e representantes de todas as equipes que criam na nuvem. Outros membros podem incluir proprietários de unidades de negócios, como marketing, que influenciarão o uso de produtos, e terceiros, como consultores para alcançar o alinhamento com seus objetivos e mecanismos e para auxiliar na geração de relatórios.
- Defina tópicos para discussão: defina os tópicos que são comuns entre as equipes ou que precisarão de um entendimento compartilhado. Siga o custo a partir do momento em que ele é criado até que a fatura seja paga. Observe todos os membros envolvidos e os processos organizacionais que devem ser aplicados. Compreenda cada etapa ou processo que ele atravessa e as informações associadas, como modelos de preços disponíveis, preços em camadas, modelos de desconto, orçamento e requisitos financeiros.

- Estabeleça uma cadência regular: para criar uma parceria financeira e tecnológica, estabeleça uma comunicação regular para criar e manter o alinhamento. O grupo precisa se reunir regularmente para comparar metas e métricas. Um ritmo típico envolve analisar o estado da organização, todos os programas em execução no momento e as métricas financeiras e de otimização gerais. Em seguida, as workloads principais são relatadas em mais detalhes.

Recursos

Documentos relacionados:

- [Notícias do blog da AWS](#)

COST01-BP03 Estabelecer orçamentos e previsões para a nuvem

Ajuste os processos de previsão e orçamento organizacional existentes para que sejam compatíveis com a natureza altamente variável dos custos e uso da nuvem. Os processos devem ser dinâmicos, usando algoritmos baseados em tendências ou em direcionadores de negócios, ou uma combinação de ambos.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Nas configurações tradicionais de TI on-premises, os clientes geralmente enfrentam o desafio de planejar custos fixos que mudam apenas ocasionalmente, em geral com a compra de novos serviços e hardware de TI para atender aos picos de demanda. Em contrapartida, a Nuvem AWS adota uma abordagem diferente na qual os clientes pagam pelos recursos que usam, conforme determinado pelas reais necessidades comerciais e de TI. No ambiente de nuvem, a demanda pode variar mensalmente, diariamente ou até mesmo de hora em hora.

O uso da nuvem traz eficiência, velocidade e agilidade, o que ocasiona um padrão de custo e de uso altamente variável. Os custos podem diminuir ou, às vezes, aumentar em resposta à maior eficiência da workload ou à implantação de novas workloads e de recursos. À medida que as workloads escalam para atender a uma base de clientes em expansão, o uso e os custos da nuvem aumentam em igual proporção em virtude da maior acessibilidade dos recursos. Essa flexibilidade nos serviços de nuvem estende-se aos custos e às previsões, criando um grau de elasticidade.

É essencial haver um estreito alinhamento com essas mudanças nas necessidades comerciais e nos motivadores de demanda e buscar o planejamento mais preciso possível. Os processos de orçamento organizacional tradicionais precisam se adaptar para acomodar essa variabilidade.

Pense na modelagem de custos ao prever o custo de novas workloads. A modelagem de custos cria uma compreensão básica dos custos esperados da nuvem, o que ajuda você a determinar o custo total de propriedade (TCO), o retorno sobre o investimento (ROI) e outras análises financeiras, definir metas e expectativas com as partes interessadas e identificar oportunidades de otimização de custos.

Sua organização precisa entender as definições de custo e os agrupamentos aceitos. O nível de detalhe no qual você prevê pode variar com base na estrutura e nos fluxos de trabalho internos da organização. Selecione um nível de granularidade que atenda aos requisitos específicos e à configuração organizacional. É importante entender em que nível a previsão é realizada:

- **Nível de conta de gerenciamento ou AWS Organizations:** a conta de gerenciamento é a conta que você usa para criar AWS Organizations. Por padrão, as organizações têm uma conta de gerenciamento.
- **Conta-membro ou vinculada:** uma conta no Organizations é uma Conta da AWS padrão que contém os seus recursos da AWS e as identidades que podem acessar esses recursos.
- **Ambiente:** um ambiente é uma coleção de recursos da AWS que executam uma versão da aplicação. Um ambiente pode ser criado com várias contas-membro ou vinculadas.
- **Projeto:** um projeto é uma combinação de objetivos ou tarefas definidas a serem realizadas dentro de um período fixo. É importante pensar no ciclo de vida do projeto durante a previsão.
- **Serviços da AWS:** grupos ou categorias, como serviços de computação ou armazenamento, nos quais você pode agrupar serviços da AWS de acordo com sua previsão.
- **Agrupamento personalizado:** é possível criar grupos personalizados com base nas necessidades da sua organização, como unidades de negócios, centros de custo, equipes, tags de alocação de custos, categorias de custo, contas vinculadas ou uma combinação delas.

Identifique os motivadores empresariais que podem afetar o custo do uso e faça uma previsão para cada um deles separadamente a fim de calcular o uso esperado com antecedência. Alguns dos motivadores podem estar vinculados às equipes de TI e de produtos da organização. Outros motivadores empresariais, como eventos de marketing, promoções, expansões geográficas, fusões e aquisições, são conhecidos por seus líderes de vendas, de marketing e de negócios, e é importante colaborar e pensar também em todos esses motivadores de demanda.

É possível usar o [AWS Cost Explorer](#) para fazer previsões baseadas em tendências em um período futuro definido com base no gasto no passado. O mecanismo de previsão do AWS Cost Explorer segmenta os dados históricos com base em tipos de cobrança (por exemplo, instâncias reservadas) e usa uma combinação de machine learning e modelos baseados em regras com a finalidade de prever os gastos individualmente para todos os tipos de cobrança.

Depois de estabelecer seu processo de previsão e criar modelos, você poderá usar o [AWS Budgets](#) para definir orçamentos personalizados em um nível granular especificando o período, a recorrência ou o valor (fixo ou variável) e adicionando filtros como serviço, Região da AWS e tags. Geralmente, o orçamento é preparado para um único ano e permanece fixo, o que exige adesão estrita de todos os envolvidos. Entretanto, a previsão é mais flexível, permitindo reajustes ao longo do ano e fornecendo projeções dinâmicas em um período de um, dois ou três anos. Tanto o orçamento quanto as previsões desempenham um papel fundamental para estabelecer expectativas financeiras entre várias partes interessadas em tecnologia e negócios. A precisão da previsão e implementação também impõe responsabilidade às partes interessadas que já são diretamente responsáveis pelo custo de provisionamento, o que também pode contribuir para o reconhecimento geral de custos.

Para se informar sobre a performance dos orçamentos atuais, é possível criar e programar relatórios do AWS Budgets para serem enviados por e-mail a você e às respectivas partes interessadas regularmente. Também é possível criar alertas do AWS Budgets com base nos custos reais, cuja natureza é reativa, ou com base nos custos previstos, que oferecem tempo para mitigar possíveis excessos de custos. Você pode receber um alerta quando o custo ou o uso realmente excederem determinado nível ou se houver previsão de que eles excederão o valor orçado.

Ajuste os processos existentes de orçamento e de previsão para se tornarem mais dinâmicos por meio de algoritmos baseados em tendências (com custos históricos como entradas) e algoritmos baseados em motivadores (por exemplo, lançamentos de novos produtos, expansão regional ou novos ambientes para workloads), que são ideais para um ambiente de gastos dinâmico e variável. Depois de determinar sua previsão baseada em tendências usando o Explorador de Custos ou qualquer outra ferramenta, use o [AWS Pricing Calculator](#) para estimar seu caso de uso da AWS e os custos futuros com base no uso esperado (tráfego, solicitações por segundo ou instâncias necessárias do Amazon EC2).

Monitore o quanto precisa é essa previsão, pois os orçamentos devem ser definidos com base nesses cálculos e estimativas. Monitore a precisão e eficácia das previsões de custos de nuvem integradas. Analise regularmente os gastos reais em comparação com a previsão e ajuste conforme necessário para torná-la mais precisa. Monitore a variação da previsão e analise a causa-raiz da variação relatada para agir e ajustar as previsões.

Conforme mencionado em [COST01-BP02 Estabelecer uma parceria entre finanças e tecnologia](#), é importante estimular parceria e ritmo entre TI, departamento financeiro e outras partes interessadas para verificar se todos usam as mesmas ferramentas e processos para manter a consistência. Nas situações em que os orçamentos precisam sofrer alterações, aumentar o ritmo dos pontos de contato pode ajudar na hora de reagir a essas mudanças com maior rapidez.

Etapas de implementação

- Defina a linguagem de custo dentro da organização: Crie uma linguagem de custo comum da AWS dentro da organização com várias dimensões e agrupamentos. Verifique se as partes interessadas entendem a granularidade das previsões, os modelos de preços e o nível de suas previsões de custos.
- Analise previsões baseadas em tendências: use ferramentas de previsão baseadas em tendências, como o AWS Cost Explorer e o Amazon Forecast. Analise o custo de uso em diferentes dimensões, como serviço, conta, tags e categorias de custos.
- Previsões baseadas em motivadores: identifique o impacto dos motivadores empresariais no uso da nuvem e faça uma previsão para cada um deles separadamente a fim de calcular o custo de uso esperado com antecedência. Trabalhe em estreita colaboração com proprietários de unidades de negócios e partes interessadas a fim de entender o impacto sobre os novos motivadores e calcular as mudanças de custo esperadas para definir orçamentos precisos.
- Atualize os processos existentes de previsão e orçamento: usando como referência métodos de previsão adotados, como baseados em tendências, baseados em motivadores de negócios ou uma combinação de ambos os métodos de previsão, defina seus processos de previsão e orçamento. Os orçamentos devem ser calculados, realistas e baseados em suas previsões.
- Configure alertas e notificações: use alertas do AWS Budgets e a detecção de anomalias em custos para receber alertas e notificações.
- Faça revisões periódicas com as principais partes interessadas: por exemplo, alinhe-se em relação às mudanças na direção e no uso dos negócios com as partes interessadas em TI, finanças, equipes de plataforma e outras áreas de negócios.

Recursos

Documentos relacionados:

- [AWS Cost Explorer](#)
- [AWS Cost and Usage Report](#)

- [Fazer previsões com o Explorador de Custos](#)
- [Fazer previsões com o Amazon QuickSight](#)
- [AWS Budgets](#)

Vídeos relacionados:

- [Como posso usar o AWS Budgets para monitorar meus gastos e uso?](#)
- [Série de otimização de custos da AWS: AWS Budgets](#)

Exemplos relacionados:

- [Entender e criar previsões baseadas em motivadores](#)
- [Como estabelecer e impulsionar uma cultura de previsão](#)
- [Como melhorar sua previsão de custos na nuvem](#)
- [Usar as ferramentas certas para prever custos na nuvem](#)

COST01-BP04 Implementar a conscientização de custos em seus processos organizacionais

Implemente a conscientização de custos, crie transparência e contabilize os custos em processos novos ou existentes que afetem o uso e aproveite os processos existentes para conscientização de custos. Implemente a conscientização de custos no treinamento de funcionários.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

A conscientização de custos deve ser implementado em processos organizacionais novos e existentes. Trata-se de um dos recursos fundamentais para outras práticas recomendadas. Recomendamos reutilizar e modificar processos existentes sempre que possível, o que minimiza o impacto na agilidade e velocidade. Informe os custos da nuvem para as equipes de tecnologia e os responsáveis por decisões nas equipes financeira e comercial para conscientizar sobre os custos, e estabeleça indicadores-chave de performance (KPIs) para as partes interessadas dos departamentos financeiro e comercial. As recomendações a seguir ajudarão a implementar a conscientização de custos em sua workload:

- Verifique se o gerenciamento de alterações inclui uma medição de custo para quantificar o impacto financeiro das mudanças. Isso ajuda a abordar de forma proativa as preocupações relacionadas a custos e a destacar as reduções de custos.
- Verifique se a otimização de custos é um componente essencial de seus recursos operacionais. Por exemplo, você pode aproveitar os processos existentes de gerenciamento de incidentes para investigar e identificar causas-raiz das anomalias de custo e uso ou excessos de custo.
- Acelere a redução de custos e a obtenção de valor empresarial por meio da automação ou de ferramentas. Ao pensar sobre o custo da implementação, enquadre a conversa para incluir um componente de retorno sobre o investimento (ROI) para justificar o investimento de tempo ou dinheiro.
- Aloque os custos de nuvem implementando showbacks ou chargebacks de gastos na nuvem, incluindo gastos com opções de compra baseadas em compromissos, serviços compartilhados e compras de marketplace para impulsionar um consumo da nuvem mais consciente sobre custos.
- Estenda os programas de treinamento e desenvolvimento existentes para incluir treinamento com conscientização de custos em toda a organização. Recomendamos que isso inclua treinamento e certificação contínuos. Isso criará uma organização capaz de autogerenciar custos e uso.
- Aproveite ao máximo as ferramentas nativas e gratuitas da AWS, como [AWS Cost Anomaly Detection](#), [AWS Budgets](#) e [AWS Budgets Reports](#).

Quando as organizações adotam consistentemente as práticas de [gerenciamento financeiro na nuvem](#) (CFM), esses comportamentos se tornam enraizados na forma de trabalhar e tomar decisões. O resultado é uma cultura mais consciente em relação aos custos, desde os desenvolvedores que arquitetam uma nova aplicação concebida na nuvem até gerentes financeiros que analisam o ROI desses novos investimentos na nuvem.

Etapas de implementação

- Identifique os processos organizacionais relevantes: cada unidade organizacional analisa os processos que possui e identifica aqueles que afetam o custo e o uso. Todos os processos que resultam na criação ou no encerramento de um recurso precisam ser incluídos para análise. Procure processos que possam sustentar a conscientização de custos na empresa, como gerenciamento de incidentes e treinamento.
- Estabeleça uma cultura com consciência de custos e autossustentável: garanta que todas as partes interessadas relevantes se alinhem ao motivo da mudança e impacto como custo para que entendam os custos da nuvem. Isso permitirá que sua organização estabeleça uma cultura de inovação autossustentável com conscientização de custos.

- Atualize os processos com conscientização de custos: cada processo é modificado para se tornar consciente dos custos. O processo pode exigir pré-verificações adicionais, como avaliação do impacto do custo, ou pós-verificações que validam se as mudanças esperadas no custo e no uso ocorreram. Processos de suporte, como treinamento e gerenciamento de incidentes, podem ser estendidos para incluir itens de custo e uso.

Para obter ajuda, fale com especialistas em CFM por meio de sua equipe de conta, ou explore os recursos e os documentos relacionados abaixo.

Recursos

Documentos relacionados:

- [Gerenciamento financeiro na Nuvem AWS](#)

Exemplos relacionados:

- [Estratégia para o gerenciamento eficiente dos custos na nuvem](#)
- [Série de blogs sobre controle de custos n.º 3: Como lidar com o impacto dos custos](#)
- [Um guia de introdução ao AWS Cost Management](#)

COST01-BP05 Relatar e notificar sobre a otimização de custos

Configure orçamentos de nuvem e mecanismos para detectar anomalias no uso. Configure ferramentas relacionadas para alertas de custo e uso em relação a metas predefinidas e receba notificações quando algum uso exceder essas metas. Faça reuniões regulares para analisar a relação custo-benefício das workloads e promover a conscientização de custos.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Informe regularmente sobre a otimização de custos e os usos dentro da sua organização. Implemente sessões dedicadas para discutir a relação de custo/performance ou inclua a otimização de custos em seus ciclos regulares de relatórios operacionais para as workloads. Use serviços e ferramentas para monitorar a relação de custo/performance regularmente e implementar oportunidades de redução de custos.

Visualize seus custos e o uso com diversos filtros e granularidades usando o [AWS Cost Explorer](#), que fornece painéis e relatórios, como custos por serviço ou por conta, custos diários ou custos de mercado. Acompanhe o andamento do custo e do uso em relação aos orçamentos configurados com o [AWS Budgets Reports](#).

Use o [AWS Budgets](#) para definir orçamentos personalizados para monitorar seus custos e uso e responder rapidamente aos alertas recebidos por e-mail ou notificações do Amazon Simple Notification Service (Amazon SNS) se você exceder seu limite. [Defina seu período orçamentário preferido](#) como diário, mensal, trimestral ou anual e crie limites orçamentários específicos para se manter informado sobre como os custos e o uso reais ou previstos progredem em relação ao limite do orçamento. Também é possível configurar [alertas](#) e [ações](#) em resposta a esses alertas para que sejam executados automaticamente ou por meio de um processo de aprovação quando uma meta de orçamento é excedida.

Implemente notificações sobre custo e uso para garantir que as mudanças no custo e no uso possam ser corrigidas rapidamente se forem inesperadas. O [AWS Cost Anomaly Detection](#) permite que você reduza as surpresas de custo e aprimore o controle sem retardar a inovação. O AWS Cost Anomaly Detection identifica gastos anômalos e causas-raiz, o que ajuda a reduzir o risco de surpresas no faturamento. Com três etapas simples, você pode criar seu próprio monitor contextualizado e receber alertas sempre que um gasto anormal for detectado.

Você também pode usar o [Amazon QuickSight](#) com dados do AWS Cost and Usage Report (CUR) para fornecer relatórios altamente personalizados com dados mais granulares. O Amazon QuickSight permite agendar relatórios e receber e-mails periódicos de relatórios de custos com informações sobre custos e usos históricos ou oportunidades de redução de custos. Confira nossa solução [Cost Intelligence Dashboard](#) (CID) criada no Amazon QuickSight para oferecer visibilidade avançada.

Use o [AWS Trusted Advisor](#), que oferece orientação para verificar se os recursos provisionados se alinham às práticas recomendadas da AWS para otimização de custos.

Verifique as recomendações de Savings Plans por meio de grafos visuais em comparação com o custo e uso detalhados. Os grafos por hora mostram os gastos sob demanda com o compromisso recomendado dos Savings Plans, fornecendo informações sobre economias estimadas, cobertura dos Savings Plans e utilização dos Savings Plans. Isso ajuda as organizações a entender como os Savings Plans se aplicam a cada hora de gasto sem precisar investir tempo e recursos na criação de modelos para analisar as despesas.

Crie periodicamente relatórios que contêm um destaque de Savings Plans, instâncias reservadas e recomendações do AWS Cost Explorer para dimensionamento do Amazon EC2 para começar a

reduzir o custo associado a workloads estacionárias e recursos ociosos ou subutilizados. Identifique e recupere os gastos associados ao desperdício de recursos implantados na nuvem. O desperdício na nuvem ocorre quando recursos dimensionados incorretamente são criados ou quando se observa padrões de uso diferentes do esperado. Siga as práticas recomendadas da AWS para reduzir o desperdício ou peça ajuda à equipe de contas e parceiro para [otimizar e economizar](#) seus gastos na nuvem.

Gere relatórios regularmente para melhorar as opções de compra de recursos a fim de reduzir os custos unitários das workloads. Opções de compra como Savings Plans, instâncias reservadas ou instâncias spot do Amazon EC2 oferecem as maiores economias para workloads tolerantes a falhas e permitem que as partes interessadas (proprietários de negócios e equipes financeiras e de tecnologia) façam parte das conversas sobre comprometimento.

Compartilhe os relatórios que contêm oportunidades ou anúncios de novos lançamentos que possam ajudar você a reduzir o custo total de propriedade (TCO) da nuvem. Adote novos serviços, regiões, recursos, soluções ou maneiras de obter mais reduções de custo.

Etapas de implementação

- Configure o AWS Budgets: configure o AWS Budgets em todas as contas para a sua workload. Defina um orçamento para o gasto total da conta e outro para a workload usando tags.
 - [Laboratórios do Well-Architected: Governança de custos e uso](#)
- Informe sobre a otimização de custos: configure um ciclo regular para discutir e analisar a eficiência da workload. Usando as métricas estabelecidas, informe sobre as métricas obtidas e o custo para alcançá-las. Identifique e corrija tendências negativas, bem como tendências positivas que possam ser promovidas em toda a organização. Os relatórios devem envolver representantes das equipes e proprietários de aplicações, bem como do setor financeiro, e os principais tomadores de decisão sobre as despesas com a nuvem.

Recursos

Documentos relacionados:

- [AWS Cost Explorer](#)
- [AWS Trusted Advisor](#)
- [AWS Budgets](#)
- [AWS Cost and Usage Report](#)

- [Práticas recomendadas do AWS Budgets](#).
- [Análise do Amazon S3](#)

Exemplos relacionados:

- [Laboratórios do Well-Architected: Governança de custos e uso](#)
- [Principais formas de começar a otimizar seus custos de nuvem da AWS](#)

COST01-BP06 Monitorar custos proativamente

Implemente ferramentas e painéis para monitorar os custos proativamente para a workload. Revise regularmente os custos com ferramentas configuradas ou prontas para usar em vez de apenas analisar os custos e as categorias quando receber notificações. O monitoramento e a análise proativa dos custos ajuda a identificar tendências positivas e permite que você as promova em toda a organização.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Recomenda-se monitorar custos e uso proativamente em sua organização, e não apenas quando há exceções ou anomalias. Painéis altamente visíveis em todo o escritório ou ambiente de trabalho garantem que as pessoas-chave tenham acesso às informações necessárias e indicam o foco da organização na otimização de custos. Os painéis visíveis permitem promover ativamente resultados bem-sucedidos e implementá-los em toda a organização.

Crie uma rotina diária ou frequente para usar o [AWS Cost Explorer](#) ou qualquer outro painel, como o [Amazon QuickSight](#), para ver os custos e fazer análises proativas. Analise o uso e os custos dos serviços da AWS na conta da AWS, no nível da workload ou em um serviço específico da AWS com agrupamento e filtragem, e valide se estão dentro do esperado ou não. Use a granularidade no nível de hora e recurso e as tags para filtrar e identificar os custos incorridos para os principais recursos. Você também pode criar seus próprios relatórios com o [Cost Intelligence Dashboard](#), uma solução do [Amazon QuickSight](#) criada por arquitetos de soluções da AWS, e comparar seus orçamentos com o custo e o uso reais.

Etapas de implementação

- Informe sobre a otimização de custos: configure um ciclo regular para discutir e analisar a eficiência da workload. Usando as métricas estabelecidas, informe sobre as métricas obtidas e o custo para alcançá-las. Identifique e corrija quaisquer tendências negativas e identifique tendências positivas a serem promovidas em toda a organização. Os relatórios devem envolver representantes das equipes de aplicações e dos proprietários, de finanças e da gerência.
- Crie e ative a granularidade diária do [AWS Budgets](#) para o custo e o uso para adotar medidas rápidas para evitar possíveis excessos de custos: o AWS Budgets permite configurar notificações de alerta para que você fique informado se algum dos seus tipos de orçamento estiver fora dos limites pré-configurados. A melhor forma de aproveitar o AWS Budgets é definir o custo e o uso esperados como limites, para que qualquer coisa acima do seu orçamento seja considerada excesso.
- Crie o AWS Cost Anomaly Detection para monitor de custos: o [AWS Cost Anomaly Detection](#) usa tecnologia avançada de machine learning para identificar gastos anormais e causas-raiz para que você possa agir rapidamente. Ele permite configurar monitores de custo que definem os segmentos de gastos que você deseja avaliar (por exemplo, serviços individuais da AWS, contas-membro, tags de alocação de custo e categorias de custo) e permite que você defina quando, onde e como recebe notificações de alerta. Para cada monitor, anexe várias assinaturas de alertas para proprietários de negócios e equipes de tecnologia, incluindo um nome, um limite de impacto do custo e a frequência de alerta (alertas individuais, resumo diário, resumo semanal) para cada assinatura.
- Use o AWS Cost Explorer ou integre seus dados do AWS Cost and Usage Report (CUR) aos painéis do Amazon QuickSight para visualizar os custos da sua organização: o AWS Cost Explorer tem uma interface fácil de usar que permite visualizar, entender e gerenciar seus custos e uso da AWS ao longo do tempo. O [Cost Intelligence Dashboard](#) é um painel personalizável e acessível que ajuda a criar a base de sua própria ferramenta de gerenciamento e otimização dos custos.

Recursos

Documentos relacionados:

- [AWS Budgets](#)
- [AWS Cost Explorer](#)
- [Orçamentos de custos e uso diários](#)
- [AWS Cost Anomaly Detection](#)

Exemplos relacionados:

- [Laboratórios do Well-Architected: Visualização](#)
- [Laboratórios do Well-Architected: Visualização avançada](#)
- [Laboratórios do Well-Architected: Cloud Intelligence Dashboards](#)
- [Laboratórios do Well-Architected: Visualização de custos](#)
- [Alerta do AWS Cost Anomaly Detection com Slack](#)

COST01-BP07 Manter-se em dia com os novos lançamentos de serviços

Consulte regularmente especialistas ou parceiros da AWS para considerar quais serviços e recursos oferecem menor custo. Revise os blogs da AWS e outras fontes de informação.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A AWS adiciona novos recursos constantemente para que você possa aproveitar as tecnologias mais recentes a fim de experimentar e inovar com maior rapidez. Você pode implementar novos serviços e recursos da AWS para aumentar a eficiência de custos na workload. Revise regularmente o [Gerenciamento de Custos da AWS](#), o [Blog de notícias da AWS](#), o [Blog de gerenciamento de custos da AWS](#) e as [Novidades da AWS](#) para obter informações sobre lançamentos de novos serviços e recursos. As postagens Novidades oferecem uma breve visão geral de todos os anúncios de serviços, recursos e expansões de regiões da AWS à medida que são lançados.

Etapas de implementação

- Inscreva-se em blogs: acesse as páginas de blogs da AWS e inscreva-se em Novidades e em outros blogs relevantes. Você pode se inscrever na página de [preferências de comunicação](#) com seu endereço de e-mail.
- Inscreva-se em Notícias da AWS: revise regularmente o [Blog de notícias da AWS](#) e [Novidades da AWS](#) para obter informações sobre novos lançamentos de serviços e recursos. Assine o feed RSS, ou use seu e-mail para ficar por dentro dos anúncios e lançamentos.
- Acompanhe as reduções de preços da AWS: cortes regulares nos preços de todos os nossos serviços são uma prática padrão que a AWS usa para passar os benefícios econômicos obtidos

pela nossa escala aos clientes. Em 20 de setembro de 2023, a AWS já havia reduzido os preços 134 vezes desde 2006. Se você tiver qualquer decisão comercial pendente por motivos de preço, poderá reavaliá-la depois das reduções de preços e das novas integrações de serviços. Aprenda sobre os esforços anteriores de redução de preços, incluindo instâncias do Amazon Elastic Compute Cloud (Amazon EC2), na [categoria de redução de preços do Blog de notícias da AWS](#).

- Eventos e reuniões da AWS: participe da conferência local da AWS e de qualquer reunião local com outras organizações da área. Se não puder participar presencialmente, tente participar dos eventos virtuais para ouvir mais de especialistas da AWS e casos de negócios de outros clientes.
- Reuna-se com sua equipe de conta: programe um ritmo regular com a equipe de conta, encontre-se com ela e discuta as tendências do setor e os serviços da AWS. Fale com seu gerente de conta, o arquiteto de soluções e a equipe de suporte.

Recursos

Documentos relacionados:

- [Gerenciamento de custos na AWS](#)
- [Novidades da AWS](#)
- [Notícias do blog da AWS](#)

Exemplos relacionados:

- [Amazon EC2: 15 anos otimizando e reduzindo custos de TI](#)
- [Blog de notícias da AWS: redução de preços](#)

COST01-BP08 Criar uma cultura de conscientização de custos

Implemente mudanças ou programas em toda a organização para criar uma cultura de conscientização de custos. Recomenda-se começar aos poucos e, à medida que seus recursos aumentarem e o uso da nuvem por sua organização crescer, implementar programas grandes e abrangentes.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Uma cultura de conscientização de custos permite escalar a otimização de custos e o gerenciamento financeiro na nuvem (operações financeiras, centro de excelência da nuvem, equipes de operações na nuvem e assim por diante) por meio de práticas recomendadas executadas de maneira orgânica e descentralizada em toda a organização. A conscientização de custos permite que você crie altos níveis de capacidade em toda a organização com o mínimo de esforço em comparação com uma abordagem centralizada e de cima para baixo.

Provocar a conscientização de custos em computação em nuvem, principalmente para geradores de custos primários na computação em nuvem, permite que as equipes entendam os resultados esperados de quaisquer alterações na perspectiva de custos. As equipes que acessam os ambientes de nuvem devem conhecer os modelos de preços e a diferença entre datacenters on-premises tradicionais e computação em nuvem.

O principal benefício de uma cultura de conscientização de custos é que as equipes de tecnologia otimizam os custos de maneira proativa e contínua (por exemplo, são consideradas um requisito não funcional ao arquitetar novas workloads ou alterar workloads existentes) em vez de realizarem otimizações de custo reativas somente quando necessárias.

Pequenas mudanças na cultura podem ter grandes impactos na eficiência de suas workloads atuais e futuras. Exemplos incluem:

- Oferecer visibilidade e conscientizar as equipes de engenharia para que entendam o que fazem e qual seu impacto em termos de custo.
- Gamificação do custo e uso em toda a organização. Isso pode ser feito por meio de um painel visível publicamente ou de um relatório que compara custos e uso normalizados entre equipes (por exemplo, custo por workload e custo por transação).
- Reconhecimento da eficiência de custos. Recompense realizações de otimização de custos voluntárias ou não solicitadas publicamente ou de forma privada e aprenda com os erros para evitar repeti-los no futuro.
- Criar requisitos organizacionais de cima para baixo para workloads a serem executadas em orçamentos predefinidos.
- Questionar os requisitos comerciais das mudanças e o impacto sobre os custos das mudanças solicitadas na infraestrutura de arquitetura ou configuração de workload para garantir que você pague somente o necessário.

- Garantir que o planejador das mudanças esteja ciente das mudanças esperadas que impactam o custo e que elas sejam confirmadas pelas partes interessadas para que proporcionem resultados comerciais com economia.

Etapas de implementação

- Informe os custos da nuvem às equipes de tecnologia: para conscientizar sobre os custos e estabelecer KPIs de eficiência para partes interessadas financeiras e comerciais.
- Informe as partes interessadas ou membros da equipe sobre mudanças planejadas: crie um item na agenda para discutir mudanças planejadas e o impacto de custo-benefício sobre a workload durante as reuniões semanais de mudanças.
- Reuna-se com sua equipe de conta: defina um cronograma de reuniões regulares com a equipe de conta, encontre-se com ela e discuta as tendências do setor e os serviços da AWS. Fale com o gerente de contas, o arquiteto e a equipe de suporte.
- Compartilhe histórias de sucesso: compartilhe histórias de sucesso sobre redução de custos de qualquer workload, Conta da AWS ou organização para gerar uma atitude positiva e encorajar sobre a otimização dos custos.
- Treinamento: garanta que as equipes técnicas ou os membros da equipe sejam treinados reconhecer os custos dos recursos na Nuvem AWS.
- Eventos e reuniões da AWS: participe de conferências locais da AWS e de quaisquer reuniões locais com outras organizações da área.
- Inscreva-se em blogs: acesse as páginas dos blogs da AWS e assine o blog [Novidades](#) e outros blogs relevantes para acompanhar os novos lançamentos, implementações, exemplos e mudanças compartilhados pela AWS.

Recursos

Documentos relacionados:

- [Blog da AWS](#)
- [Gerenciamento de custos na AWS](#)
- [Notícias do blog da AWS](#)

Exemplos relacionados:

- [Gerenciamento financeiro na Nuvem AWS](#)
- [Laboratórios do AWS Well-Architected: Gerenciamento financeiro na nuvem](#)

COST01-BP09 Quantificar o valor comercial proveniente da otimização de custos

A quantificação do valor empresarial da otimização de custos permite que você entenda todo o conjunto de benefícios da sua organização. Como a otimização de custos é um investimento necessário, quantificar o valor empresarial permite que você explique o retorno sobre o investimento para as partes interessadas. A quantificação do valor empresarial pode ajudar você a ganhar mais participação das partes interessadas em futuros investimentos de otimização de custos e fornece uma estrutura para medir os resultados das atividades de otimização de custos da sua organização.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Quantificar o valor comercial significa avaliar os benefícios que as empresas obtêm com as ações e decisões que tomam. O valor comercial pode ser tangível (como a redução das despesas ou o aumento dos lucros) ou intangível (como a melhoria da reputação da marca ou o aumento da satisfação do cliente).

Quantificar o valor comercial proveniente da otimização de custos significa determinar o valor ou o benefício que você está obtendo de seus esforços para gastar com maior eficiência. Por exemplo, se uma empresa gastar USD 100 mil para implantar uma workload na AWS e depois otimizá-la, o novo custo se tornará apenas USD 80 mil, sem prejudicar a qualidade ou a produção. Nesse cenário, o valor comercial quantificado da otimização de custos seria uma economia de USD 20 mil. Mas, além das economias, a empresa também pode quantificar o valor em termos de prazos de entrega mais rápidos, maior satisfação do cliente ou outras métricas resultantes das iniciativas de otimização de custos. As partes interessadas precisam tomar decisões sobre o valor potencial da otimização de custos, o custo da otimização da workload e o valor de retorno.

Além de relatar economias com base na otimização de custos, recomenda-se quantificar o valor adicional entregue. Os benefícios de otimização de custos normalmente são quantificados em termos de custos mais baixos por resultado comercial. Por exemplo, é possível quantificar a redução de custo do Amazon Elastic Compute Cloud (Amazon EC2) ao comprar Savings Plans, que reduzem os custos e mantêm os níveis de saída da workload. Você pode quantificar reduções de custos

nos gastos da AWS quando instâncias ociosas do Amazon EC2 são encerradas ou volumes não vinculados do Amazon Elastic Block Store (Amazon EBS) são excluídos.

No entanto, os benefícios da otimização de custos vão além da redução ou da prevenção de custos. Considere a captura de dados adicionais para medir melhorias de eficiência e valor empresarial.

Etapas de implementação

- **Avalie os benefícios comerciais:** esse é o processo de analisar e ajustar os custos da Nuvem AWS de forma a maximizar o benefício recebido de cada dólar gasto. Em vez de enfatizar a redução de custos sem valor comercial, considere os benefícios empresariais e o retorno sobre o investimento da otimização de custos, o que pode agregar maior valor ao dispêndio. Isso significa gastar com sabedoria e fazer investimentos e despesas em áreas que geram o melhor retorno.
- **Analise os custos de previsão da AWS:** a previsão ajuda as partes interessadas financeiras a definir expectativas com outras partes interessadas internas e externas da organização e pode melhorar a previsibilidade financeira da sua organização. O [AWS Cost Explorer](#) pode ser usado para realizar previsões de seu custo e uso.

Recursos

Documentos relacionados:

- [Fatores econômicos da Nuvem AWS:](#)
- [Blog da AWS](#)
- [Gerenciamento de custos na AWS](#)
- [Notícias do blog da AWS](#)
- [Whitepaper Pilar Confiabilidade do Well-Architected](#)
- [Explorador de Custos da AWS](#)

Vídeos relacionados:

- [Desbloquear o valor comercial com o Windows na AWS](#)

Exemplos relacionados:

- [Medir e maximizar o valor empresarial do Cliente 360](#)

- [O valor comercial da adoção de bancos de dados gerenciados da Amazon Web Services](#)
- [O valor comercial da Amazon Web Services para fornecedores independentes de software](#)
- [O valor empresarial da modernização na nuvem](#)
- [O valor empresarial da migração para a Amazon Web Services](#)

Reconhecimento de despesas e usos

Entender os custos e os motivadores da sua organização é essencial para gerenciar seus custos e utilização com eficácia e também identificar oportunidades de redução de custos. Normalmente, as organizações operam várias workloads executadas por várias equipes. Essas equipes podem estar em diferentes unidades da organização, cada uma com o próprio fluxo de receita. A capacidade de atribuir custos de recursos a workloads, à organização individual ou aos proprietários do produto gera um comportamento eficiente do uso e ajuda a reduzir o desperdício. O monitoramento preciso dos custos e do uso possibilita entender o quanto as unidades organizacionais e os produtos são rentáveis, permitindo que você tome decisões mais embasadas sobre onde alocar recursos dentro da sua organização. A conscientização sobre o uso em todos os níveis da organização é essencial para promover mudanças, pois a mudança no uso gera mudanças no custo.

Considere adotar uma abordagem multifacetada para se manter em dia com seus custos e utilização. Sua equipe precisa reunir dados, analisá-los e então relatá-los. Os principais fatores a serem considerados incluem:

Tópicos

- [Governança](#)
- [Monitorar custos e uso](#)
- [Desativar recursos](#)

Governança

Para gerenciar seus custos na nuvem, é necessário gerenciar seu uso por meio das seguintes áreas de governança:

Práticas recomendadas

- [COST02-BP01 Desenvolver políticas com base nos requisitos da sua organização](#)
- [COST02-BP02 Implementar metas e objetivos](#)
- [COST02-BP03 Implementar uma estrutura de contas](#)
- [COST02-BP04 Implementar grupos e perfis](#)
- [COST02-BP05 Implementar controles de custos](#)
- [COST02-BP06 rastrear o ciclo de vida do projeto](#)

COST02-BP01 Desenvolver políticas com base nos requisitos da sua organização

Desenvolva políticas que definam como os recursos são gerenciados pela sua organização e inspecione-os periodicamente. As políticas devem abranger aspectos de custos de recursos e workloads, incluindo criação, modificação e desativação ao longo da vida útil do recurso.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Entender os custos e os motivadores da sua organização é essencial para gerenciar seus custos e utilização com eficácia e também identificar oportunidades de redução de custos. Normalmente, as organizações operam várias workloads executadas por várias equipes. Essas equipes podem estar em diferentes unidades da organização, cada uma com o próprio fluxo de receita. A capacidade de atribuir os custos dos recursos a workloads, à uma organização individual ou aos donos do produto propicia um comportamento eficiente de uso e ajuda a reduzir o desperdício. O monitoramento preciso de custos e uso ajuda você a entender a otimização de sua workload, bem como a lucratividade das unidades e produtos da organização. Esse conhecimento permite uma tomada de decisão mais consciente sobre onde alocar recursos em sua organização. A conscientização sobre o uso em todos os níveis da organização é essencial para promover mudanças, pois a mudança no uso gera mudanças no custo. Considere adotar uma abordagem multifacetada para se manter em dia com seus custos e utilização.

O primeiro passo para realizar governança é usar os requisitos da sua organização para desenvolver políticas para o uso da nuvem. Essas políticas definem como a sua organização usa a nuvem e como os recursos são gerenciados. As políticas devem abranger todos os aspectos de recursos e workloads relacionados ao custo ou à utilização, incluindo criação, modificação e desativação durante a vida útil do recurso. Verifique se as políticas e os procedimentos são seguidos e implementados para qualquer alteração em um ambiente de nuvem. Durante as reuniões de gerenciamento de mudanças de TI, questione para descobrir o impacto do custo das alterações planejadas, sejam de aumento ou redução, a justificativa de negócios e o resultado esperado.

As políticas devem ser simples para que sejam facilmente compreendidas e possam ser implementadas com eficácia em toda a organização. As políticas também precisam ser fáceis de seguir e interpretar (para que sejam usadas) e específicas (para evitar erros de interpretação entre as equipes). Além disso, elas precisam ser inspecionadas periodicamente (como nossos

mecanismos) e atualizadas à medida que as condições ou as prioridades de negócios dos clientes mudam, o que tornaria a política desatualizada.

Comece com políticas amplas e de alto nível, por exemplo, qual região geográfica usar ou horários do dia em que os recursos devem estar em execução. Refine gradualmente as políticas para as várias unidades organizacionais e workloads. As políticas comuns incluem quais serviços e recursos podem ser usados (por exemplo, armazenamento de dados com menor performance em ambientes de teste e desenvolvimento), quais tipos de recursos podem ser usados por diferentes grupos (por exemplo, o maior tamanho de um recurso em uma conta de desenvolvimento é médio) e por quanto tempo esses recursos ficarão em uso (se temporariamente, em curto prazo ou por um período específico).

Exemplo de política

Veja a seguir um exemplo de política que você pode revisar para criar suas próprias políticas de governança de nuvem, que enfocam a otimização de custos. Ajuste a política com base nos requisitos de sua organização e nas solicitações das partes interessadas.

- Nome da política: defina um nome de política claro, como Política de otimização de recursos e redução de custos.
- Finalidade: explique por que essa política deve ser usada e qual é o resultado esperado. O objetivo dessa política é verificar se há um custo mínimo necessário para implantar e executar a workload desejada para atender aos requisitos de negócios.
- Escopo: defina claramente quem deve usar essa política e quando ela deve ser usada, como o DevOps X Team, para usar essa política em clientes do leste dos EUA para o ambiente X (produção ou não produção).

Declaração da política

1. Selecione us-east-1 ou várias regiões do leste dos EUA com base no ambiente de sua workload e nos requisitos de negócios (desenvolvimento, teste de aceitação do usuário, pré-produção ou produção).
2. Programe instâncias do Amazon EC2 e do Amazon RDS para execução entre 6h e 20h (Horário Padrão do Leste (EST)).
3. Interrompa todas as instâncias do Amazon EC2 não utilizadas após oito horas e as instâncias do Amazon RDS não utilizadas após 24 horas de inatividade.

4. Encerre todas as instâncias do Amazon EC2 não utilizadas após 24 horas de inatividade em ambientes que não sejam de produção. Lembre o proprietário da instância do Amazon EC2 (com base em tags) de revisar suas instâncias do Amazon EC2 paradas na produção e informá-lo de que elas serão encerradas em 72 horas se não forem usadas.
5. Use família de instância e tamanho genéricos, como m5.large, e, depois, redimensione a instância com base na utilização da CPU e da memória usando o AWS Compute Optimizer.
6. Priorize o uso do ajuste de escala automático para ajustar dinamicamente o número de instâncias em execução com base no tráfego.
7. Use instâncias spot para workloads não essenciais.
8. Analise os requisitos de capacidade para comprometer Saving Plans ou instâncias reservadas para workloads previsíveis e informe a equipe de gerenciamento financeiro da nuvem.
9. Use políticas de ciclo de vida do Amazon S3 para mover dados acessados com pouca frequência para níveis de armazenamento mais baratos. Se nenhuma política de retenção for definida, use o Amazon S3 Intelligent Tiering para mover objetos automaticamente para a camada arquivada.
10. Monitore a utilização de recursos e defina alarmes para acionar eventos de ajuste de escala usando o Amazon CloudWatch.
11. Para cada Conta da AWS, use o AWS Budgets para definir orçamentos de custo e uso para sua conta com base no centro de custos e nas unidades de negócios.
12. Usar o AWS Budgets para definir orçamentos de custo e uso para sua conta pode ajudar você a controlar seus gastos e evitar contas inesperadas, proporcionando um melhor controle sobre seus custos.

Procedimento: forneça procedimentos detalhados para implementar essa política ou consulte outros documentos que descrevam como implementar cada declaração de política. Esta seção deve fornecer instruções detalhadas para a elaboração dos requisitos da política.

Para implementar essa política, você pode usar várias ferramentas de terceiros ou regras do AWS Config para conferir a conformidade com a declaração de política e acionar ações de correção automatizadas usando funções do AWS Lambda. Você também pode usar o AWS Organizations para aplicar a política. Além disso, você deve revisar regularmente o uso de recursos e ajustar a política conforme necessário para verificar se ela continua atendendo às suas necessidades comerciais.

Etapas de implementação

- Reuna-se com as partes interessadas: para desenvolver políticas, peça às partes interessadas (escritórios de negócios na nuvem, engenheiros ou tomadores de decisão funcionais para aplicação de políticas) em sua organização que especifiquem seus requisitos e os documentem. Adote uma abordagem iterativa iniciando uma refinação ampla e contínua para as menores unidades em cada etapa. Os membros da equipe incluem aqueles com interesse direto na workload, como unidades da organização ou proprietários de aplicações, além de grupos de apoio como equipes de segurança e finanças.
- Obtenha confirmação: garanta que as equipes concordem com as políticas que determinam quem pode acessar e implantar na Nuvem AWS. Certifique-se de que elas sigam as políticas da sua organização e confirme se o provisionamento de recursos está alinhado com as políticas e procedimentos estabelecidos.
- Crie sessões de treinamento para integração: peça que os novos membros completem cursos de formação de integração para criar conscientização de custo e requisitos da organização. As equipes podem assumir políticas diferentes das suas experiências anteriores ou simplesmente ignorá-las.
- Defina locais para sua workload: defina onde sua workload opera, incluindo o país e a área dentro do país. Essas informações são usadas para mapear as zonas de disponibilidade e Regiões da AWS.
- Defina e agrupe serviços e recursos: defina os serviços que as workloads exigem. Para cada serviço, especifique os tipos, o tamanho e o número de recursos necessários. Defina grupos para os recursos por função, como servidores de aplicações ou armazenamento de banco de dados. Os recursos podem pertencer a vários grupos.
- Defina e agrupe os usuários por função: defina os usuários que interagem com a workload, concentrando-se no que eles fazem e em como usam a workload, e não em quem são nem em suas posições na organização. Agrupe usuários ou funções semelhantes. Você pode usar as políticas gerenciadas da AWS como um guia.
- Defina as ações: usando os locais, recursos e usuários identificados anteriormente, defina as ações que são exigidas por cada um para alcançar os resultados da workload ao longo do tempo de vida (desenvolvimento, operação e desativação). Identifique as ações com base nos grupos, e não nos elementos individuais nos grupos, em cada local. Comece de forma ampla como leitura ou gravação e, em seguida, refine ações específicas para cada serviço.

- Defina o período de análise: as workloads e os requisitos organizacionais podem mudar com o tempo. Defina a programação de análise da workload para garantir que ela permaneça alinhada com as prioridades organizacionais.
- Documente as políticas: verifique se as políticas que foram definidas estão acessíveis conforme exigido pela sua organização. Essas políticas são usadas para implementar, manter e auditar o acesso de seus ambientes.

Recursos

Documentos relacionados:

- [Gerenciamento de alterações na nuvem](#)
- [Políticas gerenciadas pela AWS para funções de trabalho](#)
- [Estratégia de faturamento para várias contas da AWS](#)
- [Ações, recursos e chaves de condição para serviços da AWS](#)
- [Gerenciamento e governança da AWS](#)
- [Controlar o acesso às Regiões da AWS usando as políticas do IAM](#)
- [Regiões e AZs de infraestruturas globais](#)

Vídeos relacionados:

- [Gerenciamento e governança da AWS em grande escala](#)

Exemplos relacionados:

- [VMware: o que são políticas de nuvem?](#)

COST02-BP02 Implementar metas e objetivos

Implemente metas e objetivos de custos e uso para sua workload. As metas fornecem orientação para sua organização quanto aos resultados esperados, e os objetivos oferecem resultados mensuráveis específicos a serem alcançados para suas workloads.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Desenvolva metas e objetivos de custos e uso para a sua organização. Como uma organização em crescimento na AWS, é importante definir e monitorar metas para otimização de custos. Essas metas ou [indicadores-chave de performance \(KPIs\)](#) podem incluir itens como porcentagem de gastos sob demanda ou adoção de certos serviços otimizados, como instâncias do AWS Graviton ou tipos de volume gp3 do EBS. Estabeleça metas mensuráveis e alcançáveis para ajudar a medir as melhorias de eficiência, o que é importante para suas operações comerciais. As metas fornecem orientações e direcionamento para a sua organização quanto aos resultados esperados.

Os objetivos fornecem resultados mensuráveis específicos a serem alcançados. Em suma, uma meta é a direção que você deseja seguir e o objetivo é até que ponto nessa direção a meta deve ir e quando ela deve ser concretizada [usando a orientação do método SMART (específico, mensurável, atribuível, realista e rápido)]. Um exemplo de meta é que o uso da plataforma deve aumentar significativamente, com apenas um pequeno aumento (não linear) no custo. Um exemplo de objetivo é um aumento de 20% no uso da plataforma, com um aumento de menos de 5% nos custos. Outra meta comum é que as workloads precisam ser mais eficientes a cada seis meses. O objetivo que a acompanha seria o custo de acordo com as métricas empresariais diminuir em 5% a cada seis meses. Use as métricas corretas e defina KPIs calculados para sua organização. Você pode começar com KPIs básicos e expandir posteriormente com base nas necessidades da empresa.

Uma meta para a otimização de custos é aumentar a eficiência da workload, o que corresponde a diminuir o custo por resultado empresarial da workload ao longo do tempo. Implemente essa meta para todas as workloads e defina um objetivo, como um aumento de 5% na eficiência a cada seis meses a um ano. Na nuvem, é possível conseguir isso ao estabelecer a capacidade de otimização de custos, bem como novos lançamentos de serviços e recursos.

Os objetivos são as referências quantificáveis que você deseja alcançar para concretizar suas metas, ao passo que as referências comparam seus resultados reais com um objetivo. Estabeleça referências com KPIs para o custo por unidade de serviços de computação (como adoção de spot, adoção do Graviton, tipos de instância mais recentes e cobertura sob demanda), de serviços de armazenamento (como adoção do EBS GP3, snapshots obsoletos do EBS e armazenamento do Amazon S3 padrão) ou de uso de serviços de banco de dados (como mecanismos de código aberto do RDS, adoção do Graviton e cobertura sob demanda). Essas referências e KPIs podem ajudar a verificar se você usa os serviços da AWS da maneira mais econômica.

A tabela a seguir fornece uma lista de métricas padrão da AWS para referência. Cada organização pode ter valores de objetivo diferentes para esses KPIs.

Categoria	KPI	Descrição
Computação	Cobertura de uso do EC2	Instâncias do EC2 (em custo ou horas) usando SP+RI +spot em comparação com o total (em custo ou horas) de instâncias do EC2
Computação	Utilização de SP/RI de computação	Horas de SP ou RI utilizadas em comparação com o total de horas de SP ou RI disponíveis
Computação	Custo do EC2/hora	Custo do EC2 dividido pelo número de instâncias do EC2 em execução naquela hora
Computação	Custo de vCPU	Custo por vCPU para todas as instâncias
Computação	Última geração de instância	Porcentagem de instâncias no Graviton (ou em outros tipos de instância de geração moderna)
Banco de dados	Cobertura de RDS	Instâncias do RDS (em custo ou horas) usando RI em comparação com o total (em custo ou horas) de instâncias do RDS
Banco de dados	Utilização do RDS	Horas de RI utilizadas em comparação com o total de horas de RI disponíveis
Banco de dados	Tempo de atividade do RDS	Custo do RDS dividido pelo número de instâncias do RDS em execução naquela hora

Categoria	KPI	Descrição
Banco de dados	Última geração de instância	Porcentagem de instâncias no Graviton (ou em outros tipos de instância moderna)
Armazenamento	Utilização de armazenamento	Custo do armazenamento otimizado (por exemplo, Glacier, arquivamento profundo ou acesso infrequente) dividido pelo custo total de armazenamento
Tags	Recursos não marcados	<p>Explorador de Custos</p> <ol style="list-style-type: none"> 1. Filtre créditos, descontos, impostos, reembolsos, marketplace e copie o custo mensal mais recente. 2. Selecione Mostrar somente recursos não marcados no Explorador de Custos 3. Divida o valor em recursos não marcados com seu custo mensal.

Usando essa tabela, inclua valores de metas ou de referência, os quais devem ser calculados com base nas metas organizacionais. É necessário avaliar determinadas métricas para sua empresa e entender os resultados comerciais dessa workload para definir KPIs precisos e realistas. Ao avaliar as métricas de performance em uma organização, faça a distinção entre os diferentes tipos de métrica que servem a propósitos distintos. Essas métricas avaliam principalmente a performance e a eficiência da infraestrutura técnica, e não diretamente o impacto geral nos negócios. Por exemplo, elas podem monitorar os tempos de resposta do servidor, a latência da rede ou o tempo de atividade do sistema. Essas métricas são essenciais para avaliar a capacidade da infraestrutura de comportar as operações técnicas da organização. No entanto, elas não fornecem

informações diretas sobre objetivos comerciais mais amplos, como satisfação do cliente, crescimento da receita ou participação de mercado. Para obter uma compreensão abrangente da performance dos negócios, complemente essas métricas de eficiência com métricas estratégicas de negócios que se correlacionem diretamente com os resultados comerciais.

Crie visibilidade quase em tempo real sobre seus KPIs e oportunidades de economia relacionadas e acompanhe seu progresso ao longo do tempo. Para começar a definir e monitorar os objetivos de KPI, recomendamos o painel de KPI dos [Cloud Intelligence Dashboards](#) (CID). Com base nos dados do Relatório de Custos e Uso (CUR), o painel de KPI oferece uma série de KPIs de otimização de custos recomendados com a capacidade de definir metas personalizadas e rastrear o progresso ao longo do tempo.

Se você tiver outras soluções que definam e monitorem objetivos de KPI, garanta que esses métodos sejam adotados por todas as partes interessadas de gerenciamento financeiro na nuvem em sua organização.

Etapas de implementação

- Defina os níveis de uso esperados: para começar, concentre-se nos níveis de uso. Interaja com os proprietários de aplicações, a equipe de marketing e as equipes de negócios maiores para entender quais serão os níveis de uso esperados para a workload. Como a demanda do cliente pode mudar com o tempo e o que pode mudar em decorrência de aumentos sazonais ou campanhas de marketing?
- Defina recursos e custos da workload: com os níveis de uso definidos, quantifique as alterações nos recursos da workload necessárias para atender a esses níveis de uso. Talvez seja necessário aumentar o tamanho ou o número de recursos para um componente de workload, aumentar a transferência de dados ou alterar componentes de workload para um serviço diferente em um nível específico. Especifique os custos em cada um desses pontos principais e preveja a mudança no custo quando houver uma mudança no uso.
- Defina metas de negócios: combine o resultado das alterações esperadas no uso e no custo com as alterações esperadas na tecnologia ou qualquer programa que você esteja executando e desenvolva metas para a workload. As metas devem abordar o uso e o custo, bem como a relação entre os dois. As metas devem ser simples e gerais e ajudar as pessoas a entender o que a empresa espera em termos de resultados (por exemplo, garantir que recursos não utilizados sejam mantidos abaixo de determinado nível de custo). Não é necessário definir metas para cada tipo de recurso não utilizado nem definir custos que possam causar perdas em metas e objetivos. Verifique se há programas organizacionais (por exemplo, criação de recursos como treinamento e educação) se houver alterações esperadas no custo sem alterações no uso.

- Definir objetivos: para cada uma das metas definidas, especifique um objetivo mensurável. Se a meta for aumentar a eficiência na workload, o objetivo deverá quantificar a melhoria (normalmente nos resultados de negócios para cada dólar gasto) e quando ela deverá ser entregue. Por exemplo, é possível definir uma meta para minimizar o desperdício devido ao excesso de provisionamento. Com essa meta, seu objetivo pode ser que o desperdício decorrente do superprovisionamento de computação no primeiro nível de workloads de produção não exceda 10% do custo de computação do nível. Além disso, um segundo objetivo pode ser que o desperdício decorrente do provisionamento excessivo de computação no segundo nível de workloads de produção não exceda 5% do custo de computação do nível.

Recursos

Documentos relacionados:

- [Políticas gerenciadas pela AWS para funções de trabalho](#)
- [Estratégia de faturamento para várias contas da AWS](#)
- [Controlar o acesso às Regiões da AWS usando as políticas do IAM](#)
- [Objetivos do S.M.A.R.T.](#)
- [Como rastrear KPIs de otimização de custos com o painel de KPI do CID](#)

Vídeos relacionados:

- [Laboratórios do Well-Architected: Metas e objetivos \(Nível 100\)](#)

Exemplos relacionados:

- [O que é uma métrica unitária?](#)
- [Selecionar uma métrica unitária para apoiar sua empresa](#)
- [Métricas unitárias na prática: lições aprendidas](#)
- [Como as métricas unitárias ajudam a criar alinhamento entre as funções de negócios](#)
- [Laboratórios do Well-Architected: Desativar recursos \(metas e objetivos\)](#)
- [Laboratórios do Well-Architected: Tipo, tamanho e número de recursos \(metas e objetivos\)](#)

COST02-BP03 Implementar uma estrutura de contas

Implemente uma estrutura de contas que mapeie para sua organização. Isso auxilia na alocação e no gerenciamento de custos em toda a organização.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

O AWS Organizations permite criar várias Contas da AWS que podem ajudar você a gerenciar de maneira centralizada seu ambiente à medida que dimensiona suas workloads na AWS. É possível modelar sua hierarquia organizacional agrupando Contas da AWS na estrutura da unidade organizacional (UO) e criando várias Contas da AWS em cada UO. Para criar uma estrutura de contas, primeiramente, você precisa decidir qual das suas Contas da AWS será a conta de gerenciamento. Depois disso, você pode criar Contas da AWS novas ou selecionar contas existentes como contas-membro com base na estrutura de conta projetada ao seguir as [práticas recomendadas para contas de gerenciamento](#) e as [práticas recomendadas para contas-membro](#).

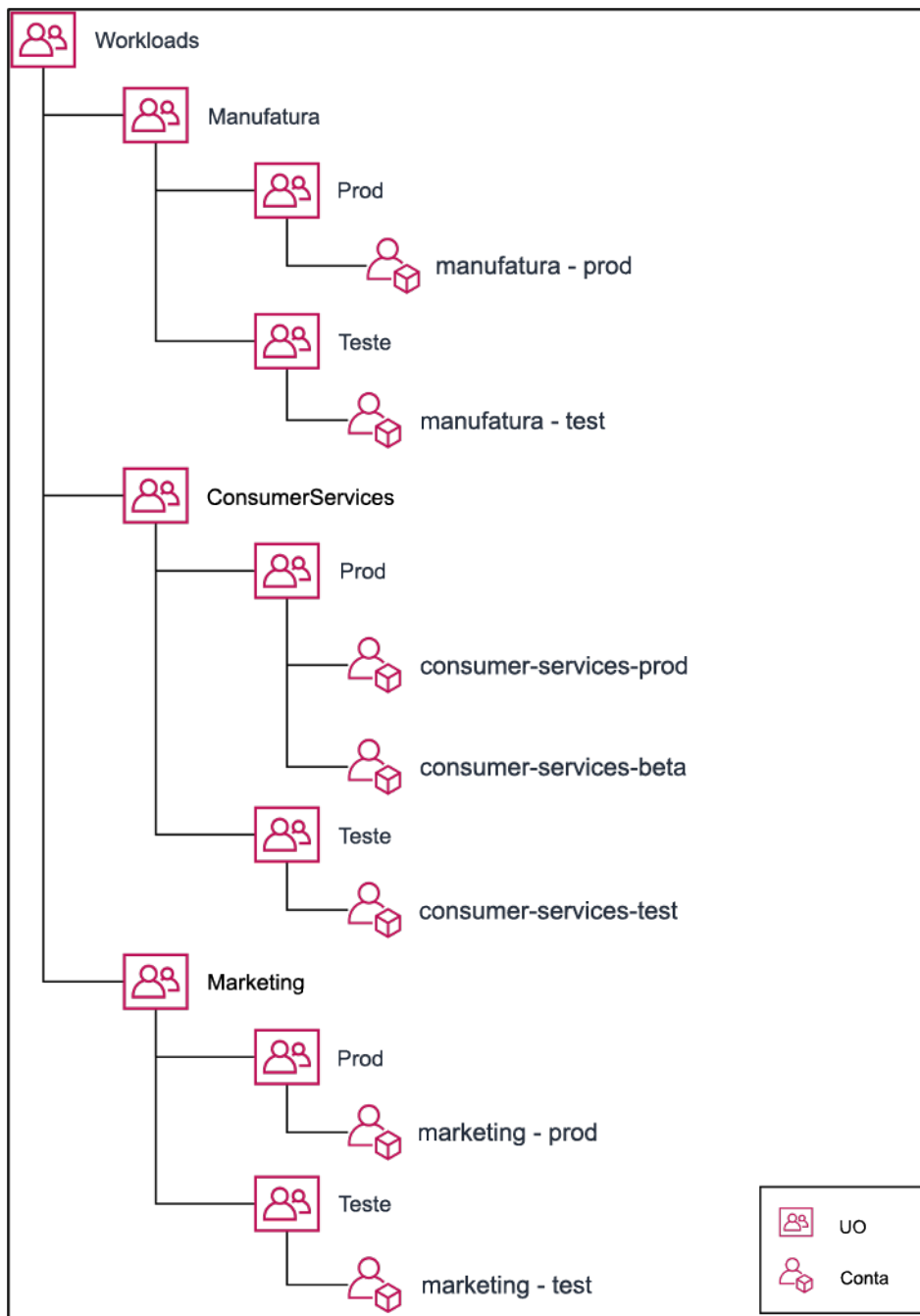
Recomenda-se que sempre haja, pelo menos, uma conta de gerenciamento com uma conta-membro, independentemente do tamanho ou uso da organização. Todos os recursos de workload devem residir somente nas contas-membro, e nenhum recurso deve ser criado na conta de gerenciamento. Não há uma resposta geral para a quantidade de Contas da AWS que você deveria ter. Avalie seus modelos de custo e operacionais atuais e futuros para garantir que a estrutura de suas Contas da AWS reflita os objetivos da sua organização. Algumas empresas criam várias Contas da AWS por motivos de negócios, por exemplo:

- O isolamento administrativo ou fiscal e de faturamento é necessário entre unidades da organização, centros de custo ou workloads específicas.
- Os limites de serviço da AWS são definidos para que sejam específicos a determinadas workloads.
- Há um requisito de isolamento e separação entre workloads e recursos.

Dentro do [AWS Organizations](#), o [faturamento consolidado](#) cria a construção entre uma ou mais contas-membro e a conta mestre. As contas-membro permitem que você isole e diferencie seu custo e uso por grupos. Uma prática comum é ter contas-membro separadas para cada unidade da organização (como finanças, marketing e vendas), ou para cada ciclo de vida do ambiente (como desenvolvimento, teste e produção) ou para cada workload (workload a, b e c) e, em seguida, agregar essas contas vinculadas usando o faturamento consolidado.

O faturamento consolidado permite consolidar o pagamento de várias Contas da AWS-membro em uma única conta de gerenciamento, sem deixar de oferecer visibilidade para a atividade de cada conta vinculada. Como os custos e o uso são agregados na conta de gerenciamento, você pode maximizar seus descontos por volume de serviço e maximizar o uso de seus descontos de compromisso (Savings Plans e instâncias reservadas) para obter os maiores descontos possíveis.

O diagrama a seguir mostra como é possível usar o AWS Organizations com unidades organizacionais (UO) para agrupar várias contas e colocar várias Contas da AWS em cada UO. Recomenda-se usar UOs para vários casos de uso e workloads que fornecem padrões para organizar contas.



Exemplo de agrupamento de várias Contas da AWS em unidades organizacionais.

O [AWS Control Tower](#) pode instalar e configurar rapidamente várias contas da AWS, garantindo que a governança esteja alinhada com os requisitos da sua organização.

Etapas de implementação

- Defina requisitos de separação: os requisitos de separação são uma combinação de vários fatores, incluindo segurança, confiabilidade e construções financeiras. Trabalhe em cada fator em ordem

e especifique se a workload ou o respectivo ambiente devem ser separados de outras workloads. A segurança promove a adesão aos requisitos de acesso e de dados. A confiabilidade gerencia os limites para que os ambientes e as workloads não afetem os outros. Revise os pilares de segurança e de confiabilidade do Well-Architected Framework periodicamente e siga as práticas recomendadas fornecidas. As estruturas financeiras criam separação financeira rígida (diferentes centros de custo, propriedades de workload e responsabilidades). Exemplos comuns de separação são workloads de produção e de teste executadas em contas separadas ou o uso de uma conta separada para que os dados da fatura e do faturamento possam ser fornecidos às unidades de negócios individuais ou aos departamentos da organização, ou à parte interessada que possui a conta.

- Defina requisitos de agrupamento: os requisitos de agrupamento não modificam os requisitos de separação, mas são usados para auxiliar o gerenciamento. Agrupe ambientes semelhantes ou workloads que não exigem separação. Um exemplo disso é o agrupamento de vários ambientes de teste ou desenvolvimento de uma ou mais workloads.
- Defina a estrutura da conta: usando essas separações e agrupamentos, especifique uma conta para cada grupo e mantenha os requisitos de separação. Essas contas são suas contas-membro ou vinculadas. Ao agrupar essas contas-membro em uma única conta de gerenciamento ou pagante, você combina o uso, o que permite maiores descontos por volume em todas as contas e fornece uma única fatura para todas as contas. É possível separar dados de faturamento e fornecer a cada conta-membro uma visualização individual dos dados de faturamento. Se uma conta-membro não precisar ter os dados de uso ou de faturamento visíveis para nenhuma outra conta, ou se uma fatura separada da AWS for necessária, você deverá definir várias contas de gerenciamento ou pagantes. Nesse caso, cada conta-membro tem a própria conta de gerenciamento ou pagante. Os recursos devem sempre ser colocados em contas-membro ou vinculadas. As contas de gerenciamento ou pagantes devem ser usadas somente para gerenciamento.

Recursos

Documentos relacionados:

- [Usar tags de alocação de custos](#)
- [Políticas gerenciadas pela AWS para funções de trabalho](#)
- [Estratégia de faturamento para várias contas da AWS](#)
- [Controlar o acesso às Regiões da AWS usando as políticas do IAM](#)
- [AWS Control Tower](#)

- [AWS Organizations](#)
- Práticas recomendadas para [contas de gerenciamento](#) e [contas-membro](#)
- [Organizar seu ambiente da AWS usando várias contas](#)
- [Ativar descontos compartilhados de instâncias reservadas e Savings Plans](#)
- [Faturamento consolidado](#)
- [Faturamento consolidado](#)

Exemplos relacionados:

- [Dividindo o CUR e compartilhando o acesso](#)

Vídeos relacionados:

- [Introdução ao AWS Organizations](#)
- [Configurar um ambiente da AWS com várias contas que use práticas recomendadas para o AWS Organizations](#)

Exemplos relacionados:

- [Laboratórios do Well-Architected: Criar um AWS Organization \(Nível 100\)](#)
- [Dividir o AWS Cost and Usage Report e compartilhar acesso](#)
- [Definir uma estratégia de várias contas da AWS para empresas de telecomunicações](#)
- [Práticas recomendadas para otimizar Contas da AWS](#)
- [Práticas recomendadas para unidades organizacionais com o AWS Organizations](#)

COST02-BP04 Implementar grupos e perfis

Implemente grupos e funções que se alinhem às políticas e controle quem pode criar, modificar ou desativar instâncias e recursos em cada grupo. Por exemplo, implemente grupos de desenvolvimento, teste e produção. Isso se aplica a serviços da AWS e a soluções de terceiros.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Os perfis e os grupos de usuários são elementos fundamentais no design e na implementação de sistemas seguros e eficientes. Os perfis e os grupos ajudam as organizações a equilibrar a necessidade de controle com a necessidade de flexibilidade e produtividade, e, acima de tudo, apoiando os objetivos organizacionais e as necessidades dos usuários. Conforme recomendado na seção [Gerenciamento de identidade e acesso](#) do Pilar Segurança do AWS Well-Architected Framework, um gerenciamento robusto de identidade e permissões é necessário para fornecer acesso aos recursos certos para as pessoas certas nas condições certas. Os usuários recebem somente o acesso necessário para realizar suas tarefas. Isso minimiza o risco associado a acesso não autorizado ou uso indevido.

Depois de desenvolver políticas, é possível criar perfis e grupos lógicos de usuários em sua organização. Isso permite que você atribua permissões, controle o uso e ajude a implementar mecanismos robustos de controle de acesso, impedindo o acesso não autorizado a informações sigilosas. Comece com agrupamentos de pessoas de alto nível. Normalmente, isso se alinha às unidades organizacionais e aos cargos (por exemplo, administrador de sistemas no departamento de TI, controlador financeiro ou analista de negócios). Os grupos categorizam pessoas que realizam tarefas semelhantes e precisam de acesso semelhante. Os perfis definem o que um grupo deve fazer. É mais fácil gerenciar permissões para grupos e perfis do que para usuários individuais. Os perfis e os grupos atribuem permissões de forma consistente e sistemática a todos os usuários, evitando erros e inconsistências.

Quando o perfil de um usuário muda, os administradores podem ajustar o acesso por perfil ou grupo, em vez de reconfigurar as contas de usuários individuais. Por exemplo, um administrador de sistemas em TI requer acesso para criar todos os recursos, mas um membro da equipe de análise só precisa criar recursos de análise.

Etapas de implementação

- Implemente grupos: usando os grupos de usuários definidos em suas políticas organizacionais, implemente os grupos correspondentes, se necessário. Para conhecer as práticas recomendadas para usuários, grupos e autenticação, consulte o [Pilar Segurança](#) do AWS Well-Architected Framework.
- Implemente perfis e políticas: usando as ações definidas em suas políticas organizacionais, crie os perfis e as políticas de acesso necessários. Para conhecer as práticas recomendadas para perfis e políticas, consulte o [Pilar Segurança](#) do AWS Well-Architected Framework.

Recursos

Documentos relacionados:

- [Políticas gerenciadas pela AWS para funções de trabalho](#)
- [Estratégia de faturamento para várias contas da AWS](#)
- [Pilar Segurança do AWS Well-Architected Framework](#)
- [AWS Identity and Access Management \(IAM\)](#)
- [Políticas do AWS Identity and Access Management](#)

Vídeos relacionados:

- [Por que usar gerenciamento de identidade e acesso](#)

Exemplos relacionados:

- [Laboratório do Well-Architected: Identidade e acesso básicos](#)
- [Controlar o acesso às Regiões da AWS usando as políticas do IAM](#)
- [Como iniciar sua jornada de gerenciamento financeiro na nuvem: operações de custos na nuvem](#)

COST02-BP05 Implementar controles de custos

Implemente controles baseados nas políticas da organização e nos perfis e grupos definidos. Isso garante que os custos sejam gerados somente conforme definido pelos requisitos da organização, como controle do acesso a regiões ou tipos de recursos.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Uma primeira etapa comum na implementação de controles de custo é configurar notificações quando eventos de custo ou de uso ocorrerem fora das políticas. É possível adotar medidas rápidas e verificar se alguma ação corretiva é necessária, sem restringir ou afetar negativamente workloads ou novas atividades. Depois de conhecer os limites da workload e do ambiente, você pode aplicar a governança. O [AWS Budgets](#) permite que você defina notificações e orçamentos mensais para seus custos, uso e descontos de compromisso (Savings Plans e Reserved Instances) da AWS. É possível criar orçamentos em um nível de custo agregado (por exemplo, todos os custos) ou em um nível

mais granular, onde você inclui apenas dimensões específicas, como contas vinculadas, serviços, tags ou zonas de disponibilidade.

Depois de definir seus limites de orçamento com o AWS Budgets, use o [AWS Cost Anomaly Detection](#) para reduzir seu custo inesperado. O AWS Cost Anomaly Detection é um serviço de gerenciamento de custos que usa machine learning para monitorar continuamente seus custos e uso para detectar gastos incomuns. Ele ajuda a identificar gastos anômalos e causas-raiz para que você possa agir rapidamente. Primeiro, crie um monitor de custos no AWS Cost Anomaly Detection e, em seguida, escolha sua preferência de alerta configurando um limite em dólares (como um alerta sobre anomalias com impacto superior a USD 1 mil). Ao receber um alerta, você poderá analisar a causa-raiz por trás da anomalia e o impacto em seus custos. Também é possível monitorar e realizar sua própria análise de anomalias no AWS Cost Explorer.

Imponha políticas de governança na AWS por meio do [AWS Identity and Access Management](#) e de [políticas de controle de serviços \(SCP\) do AWS Organizations](#). O IAM permite que você gerencie com segurança o acesso aos serviços e recursos da AWS. Com o IAM, você pode controlar quem pode criar ou gerenciar recursos da AWS, os tipos de recursos que podem ser criados e onde eles podem ser criados. Isso minimiza a possibilidade de recursos serem criados fora da política definida. Use as funções e grupos criados anteriormente e atribua [políticas do IAM](#) para impor o uso correto. Uma SCP oferece controle central sobre o número máximo de permissões disponíveis para todas as contas na sua organização, garantindo que suas contas permaneçam dentro das diretrizes de controle de acesso. As SCPs estão disponíveis somente em uma organização com todos os recursos habilitados, e você pode configurar as SCPs para negar ou permitir ações para contas-membro por padrão. Para obter mais detalhes sobre a implementação do gerenciamento de acesso, consulte o [whitepaper Pilar Segurança do Well-Architected](#).

A governança também pode ser implementada por meio do gerenciamento de [cotas de serviço da AWS](#). Ao garantir que as cotas de serviço sejam configuradas com o mínimo de sobrecarga e mantidas com precisão, você pode minimizar a criação de recursos fora dos requisitos da sua organização. Para conseguir isso, você deve entender a rapidez com que seus requisitos podem mudar, compreender projetos em andamento (criação e desativação de recursos) e considerar a rapidez com que as alterações de cota podem ser implementadas. As [cotas de serviço](#) podem ser usadas para aumentar suas cotas quando necessário.

Etapas de implementação

- Implemente notificações sobre gastos: usando suas políticas organizacionais definidas, crie [AWS Budgets](#) para receber notificações quando os gastos estiverem fora de suas políticas. Configure vários orçamentos de custos, um para cada conta, para ser notificado sobre os gastos gerais da

conta. Configure orçamentos de custos adicionais dentro de cada conta para unidades menores dentro da conta. Essas unidades variam de acordo com a estrutura da sua conta. Alguns exemplos comuns são Regiões da AWS, workloads (usando tags) ou serviços da AWS. Configure uma lista de distribuição de e-mails como o destinatário das notificações, e não uma conta de e-mail de uma pessoa. É possível configurar um orçamento real para quando um valor for ultrapassado ou usar um orçamento previsto para notificar sobre o uso previsto. Você também pode pré-configurar ações do AWS Budgets que podem aplicar políticas específicas do IAM ou SCP ou interromper instâncias do Amazon EC2 ou Amazon RDS. As ações de orçamento podem ser executadas automaticamente ou exigir aprovação do fluxo de trabalho.

- Implemente notificações sobre gastos anômalos: use o [AWS Cost Anomaly Detection](#) para reduzir custos inesperados em sua organização e analisar a causa-raiz de possíveis gastos anômalos. Depois de criar o monitor de custos para identificar gastos incomuns em sua granularidade especificada e configurar notificações no AWS Cost Anomaly Detection, ele envia um alerta quando um gasto incomum é detectado. Isso permitirá que você analise a causa-raiz por trás da anomalia e entenda o impacto em seu custo. Use Categorias de Custos da AWS durante a configuração do AWS Cost Anomaly Detection para identificar qual equipe de projeto ou equipe de unidade de negócios pode analisar a causa-raiz do custo inesperado e tomar as ações necessárias em tempo hábil.
- Implemente controles de utilização: usando as políticas da organização definidas, implemente políticas e perfis do IAM para especificar quais ações os usuários podem e não podem executar. Várias políticas organizacionais podem ser incluídas em uma política da AWS. Da mesma forma que você definiu políticas, comece de forma mais ampla e, em seguida, aplique controles mais granulares em cada etapa. Os limites de serviço também são um controle eficaz do uso. Implemente os limites de serviço corretos em todas as suas contas.

Recursos

Documentos relacionados:

- [Políticas gerenciadas pela AWS para funções de trabalho](#)
- [Estratégia de faturamento para várias contas da AWS](#)
- [Controlar o acesso às Regiões da AWS usando as políticas do IAM](#)
- [AWS Budgets](#)
- [AWS Cost Anomaly Detection](#)
- [Controlar seus custos da AWS](#)

Vídeos relacionados:

- [Como posso usar o AWS Budgets para monitorar meus gastos e uso?](#)

Exemplos relacionados:

- [Exemplos de políticas de gerenciamento de acesso do IAM](#)
- [Exemplos de políticas de controle de serviços](#)
- [AWS Budgets Actions](#)
- [Criar uma política do IAM para controlar o acesso aos recursos do Amazon EC2 usando tags](#)
- [Restringir o acesso da identidade do IAM a recursos específicos do Amazon EC2](#)
- [Criar uma política do IAM para restringir o uso do Amazon EC2 pela família](#)
- [Laboratórios do Well-Architected: Governança de custos e uso \(Nível 100\)](#)
- [Laboratórios do Well-Architected: Governança de custos e uso \(Nível 200\)](#)
- [Integrações do Slack para detecção de anomalias de custo usando o Amazon Q Developer em aplicações de chat](#)

COST02-BP06 rastrear o ciclo de vida do projeto

Acompanhe, meça e realize auditorias no ciclo de vida dos projetos, equipes e ambientes para evitar o uso e pagamento de recursos desnecessários.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Ao monitorar com eficácia o ciclo de vida do projeto, as organizações podem ter um melhor controle de custos por meio de planejamento, gerenciamento e otimização de recursos aprimorados. Os insights recebidos por meio do rastreamento são inestimáveis para a tomada de decisões fundamentadas que contribuem para a relação custo-benefício e o sucesso geral do projeto.

O rastreamento de todo o ciclo de vida da workload ajuda a compreender quando as workloads ou os respectivos componentes não são mais necessários. As workloads e os componentes existentes podem parecer estar em uso, mas quando a AWS libera novos serviços ou recursos, eles podem ser desativados ou adotados. Confira os estágios anteriores das workloads. Depois que uma

workload entra em produção, os ambientes anteriores podem ser desativados ou ter a capacidade significativamente reduzida até que sejam necessários novamente.

É possível marcar recursos com um prazo ou um lembrete para fixar a hora em que a workload foi analisada. Por exemplo, se o ambiente de desenvolvimento foi analisado pela última vez meses atrás, talvez seja um bom momento para analisá-lo novamente com o objetivo de examinar se novos serviços podem ser adotados ou se o ambiente está em uso. Você pode agrupar e marcar aplicações com o [myApplications](#) na AWS para gerenciar e rastrear metadados, como criticidade, ambiente, última revisão e centro de custos. É possível monitorar o ciclo de vida da workload e monitorar e gerenciar o custo, a integridade, o procedimento de segurança e a performance das aplicações.

A AWS fornece uma série de serviços de gerenciamento e de governança que você pode usar para o monitoramento do ciclo de vida da entidade. É possível usar o [AWS Config](#) ou o [AWS Systems Manager](#) para fornecer um inventário detalhado dos recursos e da configuração da AWS. Recomendamos integrá-lo a seus sistemas existentes de gerenciamento de projetos ou ativos para rastrear projetos e produtos ativos em sua organização. A combinação do seu sistema atual com o conjunto de eventos e métricas avançados fornecido pela AWS permite criar uma visão de eventos de ciclo de vida significativos e gerenciar os recursos proativamente para reduzir os custos desnecessários.

Semelhante ao [gerenciamento do ciclo de vida da aplicação \(ALM\)](#), o rastreamento do ciclo de vida do projeto deve envolver vários processos, ferramentas e equipes trabalhando em conjunto, como design e desenvolvimento, testes, produção, suporte e redundância de workload.

Ao monitorar cuidadosamente cada fase do ciclo de vida de um projeto, as organizações obtêm insights cruciais e um controle aprimorado, o que facilita o sucesso do planejamento, da implementação e da conclusão do projeto. Essa supervisão cuidadosa verifica se os projetos, além de atenderem aos padrões de qualidade, são entregues no prazo e dentro do orçamento, promovendo o custo-benefício de modo geral.

Consulte o [whitepaper Pilar Excelência operacional do AWS Well-Architected](#) para obter mais detalhes sobre a implementação do rastreamento do ciclo de vida da entidade.

Etapas de implementação

- Estabeleça o processo de monitoramento do ciclo de vida do projeto: a [equipe do Centro de Excelência da Nuvem](#) deve estabelecer o processo de monitoramento do ciclo de vida do projeto. Estabeleça uma abordagem estruturada e sistemática para monitorar as workloads a fim de melhorar o controle, a visibilidade e a performance dos projetos. Torne o processo de

monitoramento transparente, colaborativo e dedicado à melhoria contínua para maximizar sua eficácia e valor.

- Realize análises da workload: conforme definido por suas políticas organizacionais, configure uma frequência regular para auditar seus projetos existentes e realizar análises da workload. A quantidade de esforço empregado na auditoria deve ser proporcional ao risco aproximado, ao valor ou ao custo para a organização. As principais áreas a serem incluídas na auditoria seriam riscos para a organização de um incidente ou interrupção, valor ou contribuição para a organização (medidos em receita ou reputação da marca), custo da workload (medido como custo total de recursos e custos operacionais) e uso da workload (medido em número de resultados da organização por unidade de tempo). Se essas áreas mudarem ao longo do ciclo de vida, ajustes serão necessários na workload, como desativação total ou parcial.

Recursos

Documentos relacionados:

- [Orientação para marcação com tags na AWS](#)
- [O que é gerenciamento do ciclo de vida de aplicações \(ALM\)?](#)
- [Políticas gerenciadas pela AWS para funções de trabalho](#)

Exemplos relacionados:

- [Controlar o acesso às Regiões da AWS usando as políticas do IAM](#)

Ferramentas relacionadas

- [AWS Config](#)
- [AWS Systems Manager](#)
- [AWS Budgets](#)
- [AWS Organizations](#)
- [AWS CloudFormation](#)

Monitorar custos e uso

Permita que as equipes atuem em seus custos e uso por meio de visibilidade detalhada da workload. A otimização de custos começa com uma compreensão granular da discriminação de custos e uso, da capacidade de modelar e prever gastos, uso e recursos futuros e da implementação de mecanismos suficientes para alinhar custos e uso aos objetivos da sua organização. Veja a seguir as áreas necessárias para monitorar seus custos e uso:

Práticas recomendadas

- [COST03-BP01 Configurar fontes de informações detalhadas](#)
- [COST03-BP02 Adicionar informações da organização aos custos e ao uso](#)
- [COST03-BP03 Identificar categorias de atribuição de custos](#)
- [COST03-BP04 Estabelecer métricas da organização](#)
- [COST03-BP05 Configurar ferramentas de faturamento e gerenciamento de custos](#)
- [COST03-BP06 Alocar custos com base nas métricas de workload](#)

COST03-BP01 Configurar fontes de informações detalhadas

Configure ferramentas de gerenciamento de custos e geração de relatórios para aprimorar as análises e obter transparência dos dados de custo e uso. Configure a workload para criar entradas de log que facilitem o rastreamento e a segmentação de custos e uso.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Informações detalhadas de faturamento, como granularidade por hora nas ferramentas de gerenciamento de custos, permitem que as organizações acompanhem suas taxas de consumo com mais detalhes e as ajudem a identificar alguns dos motivos do aumento de custos. Essas fontes de dados oferecem a visualização mais precisa do custo e do uso em toda a organização.

Você pode usar o Exportações de dados da AWS para criar exportações do AWS Cost and Usage Report (CUR) 2.0. Ele é a nova forma recomendada de receber dados de custo e uso detalhados da AWS. Ele fornece granularidade de uso diário ou por hora, taxas, custos e atributos de uso para todos os serviços da AWS cobráveis (as mesmas informações do CUR), além de algumas melhorias. Todas as dimensões possíveis estão no CUR, incluindo marcação, localização, atributos de recurso e IDs de conta.

Há três tipos de exportação com base no tipo de exportação que você deseja criar: uma exportação de dados padrão, uma exportação para um painel de custos e uso integrado ao Amazon QuickSight ou uma exportação de dados herdados.

- Exportação de dados padrão: exportação personalizada de uma tabela que é entregue ao Amazon S3 de forma recorrente.
- Painel de Custos e Uso: exportação e integração com o Amazon QuickSight que implanta um painel pré-criado de custos e uso.
- Exportação de dados herdados: uma exportação do AWS Cost and Usage Report (CUR) herdado.

É possível criar exportações de dados com as seguintes personalizações:

- Incluir IDs de recurso
- Dados de alocação de custos divididos
- Detalhamento por hora
- Versionamento
- Tipo de compactação e formato de arquivo

Para workloads que executam contêineres no Amazon EC2 ou no Amazon EKS, habilite os dados de alocação de custos divididos para que você possa alocar seus custos de contêiner para unidades de negócios e equipes individuais, com base em como as workloads de contêiner consomem os recursos compartilhados de computação e memória. Os dados de alocação de custos divididos apresentam dados de custos e uso de novos recursos em nível de contêiner ao AWS Cost and Usage Report. Os dados de alocação de custos divididos são calculados computando-se os custos de serviços e tarefas individuais do ECS em execução no cluster.

Um painel de custos e uso exporta a tabela do painel de custos e uso para um bucket do S3 de forma recorrente e implanta um painel de custos e uso predefinido no Amazon QuickSight. Use essa opção se quiser implantar rapidamente um painel de seus dados de custos e uso sem a possibilidade de personalização.

Se desejar, você ainda poderá exportar o CUR no modo herdado, onde é possível integrar outros serviços de processamento, como o [AWS Glue](#), para preparar os dados para análise e realizar análises de dados com o [Amazon Athena](#) usando SQL para consultar os dados.

Etapas de implementação

- Crie exportações de dados: crie exportações personalizadas com os dados desejados e controle o esquema das suas exportações. Crie exportações de dados de gerenciamento de custos e cobrança usando SQL básico e visualizar os dados de gerenciamento de custos e faturamento por meio da integração com o Amazon QuickSight. Você também pode exportar seus dados no modo padrão para analisá-los com outras ferramentas de processamento, como o Amazon Athena.
- Configure o relatório de custos e uso: usando o console de faturamento, configure pelo menos um relatório de custos e uso. Configure um relatório com granularidade por hora que inclua todos os identificadores e IDs de recursos. Você também pode criar outros relatórios com diferentes níveis de detalhamento para fornecer informações resumidas de alto nível.
- Configure a granularidade por hora no Explorador de Custos: para acessar dados de custo e uso com granularidade horária dos últimos 14 dias, considere ativar dados horários e de nível de recursos no console de faturamento.
- Configure o log da aplicação: verifique se a aplicação registra cada resultado comercial entregue para que ele possa ser acompanhado e medido. Verifique se a granularidade desses dados é pelo menos por hora para corresponder aos dados de custo e uso. Para obter mais detalhes sobre registro e monitoramento, consulte [Pilar Excelência operacional do Well-Architected](#).

Recursos

Documentos relacionados:

- [Exportações de dados da AWS](#)
- [AWS Glue](#)
- [Amazon QuickSight](#)
- [Preços do gerenciamento de custos da AWS](#)
- [Marcando recursos do AWS](#)
- [Analisar custos com Explorador de Custos](#)
- [Gerenciar AWS Cost and Usage Reports](#)
- [Pilar Excelência operacional da Well-Architected](#)

Exemplos relacionados:

- [Configuração da conta da AWS](#)

- [Exportação de dados para o Gerenciamento de Faturamento e Custos da AWS](#)
- [Casos de uso comuns do AWS Cost Explorer](#)

COST03-BP02 Adicionar informações da organização aos custos e ao uso

Defina um esquema de marcação com tags com base na sua organização, atributos da workload e categorias de alocação de custos para que você possa filtrar e pesquisar recursos ou monitorar custos e uso em ferramentas de gerenciamento de custos. Implemente marcação consistente em todos os recursos, sempre que possível, por finalidade, equipe, ambiente ou outros critérios relevantes ao seu negócio.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Implemente a [marcação com tags na AWS](#) para adicionar informações da organização aos seus recursos, as quais serão adicionadas às suas informações de custo e uso. Uma tag é um par de chave-valor. A chave é definida e deve ser exclusiva em toda a organização, e o valor é exclusivo para um grupo de recursos. Um exemplo de par chave-valor é a chave `Environment`, com um valor de `Production`. Todos os recursos no ambiente de produção terão esse par de chave-valor. A marcação com tags permite categorizar e rastrear seus custos com informações relevantes e significativas da organização. Você pode aplicar tags que representem categorias de organização (como centros de custo, nomes de aplicações, projetos ou proprietários) e identificar workloads e características de workloads (como teste ou produção) para atribuir seus custos e uso em toda a organização.

Quando você aplica tags aos seus recursos da AWS (como instâncias do Amazon Elastic Compute Cloud ou buckets do Amazon Simple Storage Service) e as ativa, a AWS adiciona essas informações aos relatórios de custos e uso. Você pode gerar relatórios e realizar análises em recursos marcados e não marcados para permitir maior conformidade com políticas internas de gerenciamento de custos e garantir a atribuição precisa.

Criar e implementar um padrão de marcação da AWS em todas as contas da organização ajuda você a gerenciar e administrar seus ambientes da AWS de maneira consistente e uniforme. Use [políticas de tag](#) no AWS Organizations para definir regras de como as tags podem ser usadas em recursos da AWS nas suas contas no AWS Organizations. As políticas de tag permitem adotar facilmente uma abordagem padronizada para marcar os recursos da AWS.

O [AWS Tag Editor](#) permite adicionar, excluir e gerenciar tags de vários recursos. Com o Tag Editor, você pode pesquisar os recursos que deseja marcar e gerenciar as tags dos recursos nos resultados da pesquisa.

As [Categorias de Custos da AWS](#) permitem que você atribua significado da organização aos seus custos, sem exigir tags nos recursos. É possível mapear suas informações de custo e uso em estruturas internas exclusivas da organização. Você define regras de categoria para mapear e categorizar custos usando dimensões de faturamento, como contas e tags. Isso fornece outro nível de capacidade de gerenciamento, além da marcação. Você também pode mapear contas e tags específicas para vários projetos.

Etapas de implementação

- Defina um esquema de marcação com tags: reúna todas as partes interessadas de toda a sua empresa para definir um esquema. Isso geralmente inclui pessoas dos departamentos técnico, financeiro e de gerenciamento. Defina uma lista de tags que todos os recursos devem obrigatoriamente ter, bem como outra lista com as tags que os recursos poderiam ter. Verifique se os nomes e valores das tags são consistentes em toda a organização.
- Marque recursos com tags: usando suas categorias de atribuição de custo definidas, [coloque tags](#) em todos os recursos em suas workloads de acordo com as categorias. Use ferramentas como CLI, Tag Editor ou AWS Systems Manager para aumentar a eficiência.
- Implemente Categorias de Custos da AWS: você pode criar [categorias de custos](#) sem implementar a marcação com tags. As categorias de custos usam as dimensões de custo e uso existentes. Crie regras de categoria a partir do esquema e as implemente nas categorias de custos.
- Automatize a marcação com tags: para garantir que você mantenha altos níveis de marcação em todos os recursos, automatize a marcação com tags para que os recursos sejam marcados automaticamente quando forem criados. Use serviços como o [AWS CloudFormation](#) para verificar se os recursos são marcados quando criados. Você também pode criar uma solução personalizada para fazer a marcação com tags automaticamente usando funções do Lambda ou usar um microsserviço personalizado que verifica a workload periodicamente e remove todos os recursos que não estão marcados, o que é ideal para ambientes de teste e desenvolvimento.
- Monitore e gere relatórios de tags: para garantir que você mantenha altos níveis de marcação em toda a organização, relate e monitore as tags em todas as workloads. É possível usar o [AWS Cost Explorer](#) para visualizar o custo de recursos marcados e não marcados ou usar serviços como o [Tag Editor](#). Analise regularmente o número de recursos não marcados com tags e tome medidas para adicionar etiquetas até atingir o nível desejado de marcação.

Recursos

Documentos relacionados:

- [Práticas recomendadas de marcação com tags](#)
- [Tag de recurso do AWS CloudFormation](#)
- [AWS Cost Categories](#)
- [Marcando recursos do AWS](#)
- [Analisar os custos com o AWS Budgets](#)
- [Analisar custos com Explorador de Custos](#)
- [Gerenciar Relatórios de Custos e Uso da AWS](#)

Vídeos relacionados:

- [Como posso marcar meus recursos da AWS com tags para dividir minha fatura por centro de custos ou projeto?](#)
- [Marcar recursos da AWS com tags](#)

COST03-BP03 Identificar categorias de atribuição de custos

Identifique categorias organizacionais, como unidades de negócios, departamentos ou projetos, que poderiam ser usadas para alocar custos em sua organização às entidades consumidoras internas. Use essas categorias para impor a responsabilidade de gastos, bem como promover a conscientização de custos e comportamentos de consumo eficazes.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

O processo de categorização de custos é crucial em orçamentos, contabilidade, relatórios financeiros, tomada de decisão, benchmarking e gerenciamento de projetos. Ao classificar e categorizar as despesas, as equipes podem entender melhor os tipos de custos que gerarão ao longo da jornada para a nuvem, ajudando-as a tomar decisões conscientes e gerenciar orçamentos de forma eficaz.

A responsabilidade pelos gastos com a nuvem estabelece um forte incentivo para o gerenciamento disciplinado da demanda e dos custos. O resultado é uma economia significativamente maior nos

custos da nuvem para organizações que alocam a maior parte de seus gastos com a nuvem para unidades de negócios ou equipes consumidoras. Além disso, a alocação de gastos na nuvem ajuda as organizações a adotar mais práticas recomendadas de governança centralizada da nuvem.

Trabalhe com sua equipe financeira e outras partes interessadas relevantes para entender os requisitos de como os custos devem ser alocados em sua organização durante suas chamadas regulares. Os custos da workload devem ser alocados durante todo o ciclo de vida, incluindo desenvolvimento, teste, produção e desativação. Entenda como os custos incorridos para o aprendizado, o desenvolvimento da equipe e a criação de ideias são atribuídos na organização. Isso pode ser útil para alocar corretamente contas usadas para essa finalidade para orçamentos de treinamento e desenvolvimento, em vez de orçamentos genéricos de custo de TI.

Depois de definir suas categorias de atribuição de custos com as partes interessadas em sua organização, use [Categorias de Custos da AWS](#) para agrupar suas informações de custos e uso em categorias significativas na Nuvem AWS, como custo de um projeto específico, ou Contas da AWS para departamentos ou unidades de negócios. É possível criar categorias personalizadas e mapear as informações de custo e uso nessas categorias com base nas regras definidas usando várias dimensões, como conta, tag, serviço ou tipo de cobrança. Assim que as categorias de custos forem definidas, você verá as informações de custos e uso de acordo com elas, permitindo que a organização tome melhores decisões estratégicas e de compras. Também é possível ver essas categorias no AWS Cost Explorer, no AWS Budgets e no AWS Cost and Usage Report.

Por exemplo, é possível criar categorias de custos para suas unidades de negócios (equipe DevOps) e, em cada categoria, criar várias regras (para cada subcategoria) com várias dimensões (Contas da AWS, tags de alocação de custos, serviços ou tipo de cobrança) com base nos seus agrupamentos definidos. Com as categorias de custo, é possível organizar seus custos usando um mecanismo baseado em regras. As regras que você configurar organizarão seus custos em categorias. Dentro dessas regras, é possível aplicar filtros usando várias dimensões para cada categoria, como Contas da AWS, serviços da AWS ou tipos de cobrança específicos. Você pode usar essas categorias em vários produtos do [console](#) do [Gerenciamento de Faturamento e Custos da AWS Billing and Cost Management](#). Isso inclui AWS Cost Explorer, AWS Budgets, AWS Cost and Usage Report e AWS Cost Anomaly Detection.

Como exemplo, o diagrama a seguir mostra como agrupar as informações de custos e uso em sua organização, com várias equipes (categoria de custos), vários ambientes (regras) e vários recursos ou ativos em cada ambiente (dimensões).

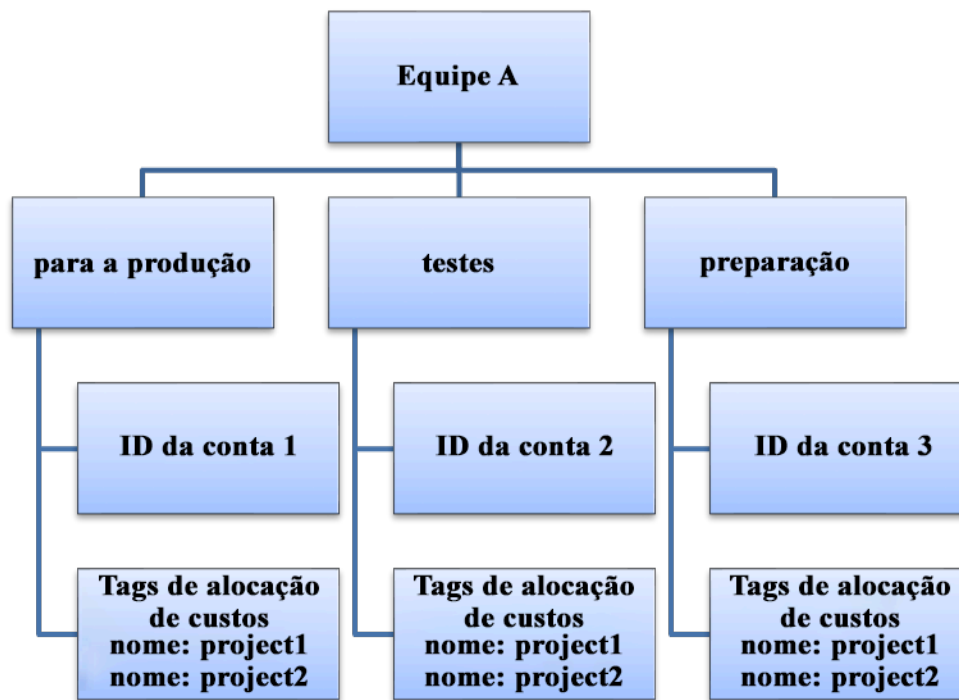


Tabela de organização de custos e uso

Também é possível criar agrupamento de custos usando as categorias de custos. Depois de criar as categorias de custos (aguardando até 24 horas após a criação de uma categoria para que seus registros de uso sejam atualizados com valores), elas aparecem no [AWS Cost Explorer](#), [AWS Budgets](#), [AWS Cost and Usage Report](#) e [AWS Cost Anomaly Detection](#). No AWS Cost Explorer e no AWS Budgets, uma categoria de custos aparece como uma dimensão de faturamento adicional. Você pode usar isso para filtrar o valor da categoria de custo específica ou agrupar pela categoria de custo.

Etapas de implementação

- Defina as categorias da sua organização: reúna-se com as unidades de negócios e as partes interessadas internas para definir categorias que reflitam a estrutura e os requisitos da organização. Essas categorias devem ser associadas diretamente à estrutura das categorias financeiras existentes, como unidade de negócios, orçamento, centro de custo ou departamento. Veja os resultados que a nuvem oferece para a sua empresa, como treinamento ou educação, já que também são categorias de organização.
- Defina suas categorias funcionais: reúna-se com as unidades de negócios e as partes interessadas internas para definir categorias que reflitam as funções presentes na empresa. Essas

podem ser os nomes da workload ou da aplicação e o tipo de ambiente, como produção, teste ou desenvolvimento.

- Defina Categorias de Custos da AWS: crie categorias de custos para organizar suas informações de custo e uso usando [Categorias de Custos da AWS](#) e mapeie seus custos e uso da AWS em [categorias significativas](#). Várias categorias podem ser atribuídas a um recurso, e um recurso pode estar em várias categorias diferentes. Portanto, defina quantas categorias forem necessárias para [gerenciar seus custos](#) dentro da estrutura categorizada usando Categorias de Custos da AWS.

Recursos

Documentos relacionados:

- [Marcando recursos do AWS](#)
- [Usar tags de alocação de custos](#)
- [Analisar custos com o AWS Budgets](#)
- [Analisar custos com Explorador de Custos](#)
- [Gerenciar AWS Cost and Usage Reports](#)
- [AWS Cost Categories](#)
- [Gerenciar seus custos com as Categorias de Custos da AWS](#)
- [Criar categorias de custos](#)
- [Marcar categorias de custos com tags](#)
- [Separar cobranças em categorias de custos](#)
- [Recursos das Categorias de Custos da AWS](#)

Exemplos relacionados:

- [Organizar seus dados de custos e uso com as Categorias de Custos da AWS](#)
- [Gerenciar seus custos com as Categorias de Custos da AWS](#)
- [Laboratórios do Well-Architected: Visualização de custos e uso](#)
- [Laboratórios do Well-Architected: Categorias de custos](#)

COST03-BP04 Estabelecer métricas da organização

Estabeleça as métricas da organização que são necessárias para esta workload. Exemplo de métricas de uma workload são relatórios de clientes produzidos ou páginas da Web veiculadas para os clientes.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Entenda como a saída da workload é medida em relação ao sucesso empresarial. Cada workload normalmente tem um pequeno conjunto de saídas principais que indicam a performance. Se você tiver uma workload complexa com muitos componentes, poderá priorizar a lista ou definir e rastrear métricas para cada componente. Trabalhe com suas equipes para entender quais métricas usar. Essa unidade será usada para compreender a eficiência da workload ou o custo de cada saída de negócios.

Etapas de implementação

- Defina resultados da workload: reúna-se com as partes interessadas da empresa e defina os resultados para a workload. Essas são medidas principais de uso do cliente e devem ser métricas de negócios, e não técnicas. Deve haver um pequeno número de métricas de alto nível (menos de cinco) por workload. Se a workload produzir vários resultados para diferentes casos de uso, agrupe-os em uma única métrica.
- Definir resultados dos componentes da workload: opcionalmente, se você tiver uma workload grande e complexa ou puder dividir facilmente sua workload em componentes (como microsserviços) com entradas e saídas bem definidas, defina métricas para cada componente. O esforço deve refletir o valor e o custo do componente. Comece com os maiores componentes e trabalhe em direção aos componentes menores.

Recursos

Documentos relacionados:

- [Marcando recursos do AWS](#)
- [Analisar os custos com o AWS Budgets](#)
- [Analisar custos com Explorador de Custos](#)
- [Gerenciar Relatórios de Custos e Uso da AWS](#)

COST03-BP05 Configurar ferramentas de faturamento e gerenciamento de custos

Configure ferramentas de gerenciamento de custos que atendam às políticas da sua organização para gerenciar e otimizar gastos com a nuvem. Isso inclui serviços, ferramentas e recursos para organizar e monitorar dados de custos e uso, aprimorar o controle por meio de faturamento consolidado e permissão de acesso, melhorar o planejamento por meio de orçamento e previsões, receber notificações ou alertas e reduzir os custos com recursos e otimizações de preços.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Para estabelecer uma responsabilização consistente, pense primeiro na estratégia de conta como parte de sua estratégia de alocação de custos. Faça isso do jeito certo e talvez não precise ir além. Caso contrário, poderá haver inconsciência e outros pontos problemáticos.

Para incentivar a responsabilização pelos gastos com a nuvem, conceda aos usuários acesso a ferramentas que forneçam visibilidade sobre custos e uso. A AWS recomenda configurar todas as workloads e as equipes para as seguintes finalidades:

- **Organize:** estabeleça a alocação de custos e linha de base de governança com sua própria estratégia de marcação e taxonomia. Crie várias contas da AWS com ferramentas como o AWS Control Tower ou o AWS Organization. Marque os recursos compatíveis da AWS com tags e categorize-os de uma maneira que faça sentido com base na estrutura da sua organização (unidades de negócios, departamentos ou projetos). Marque nomes de contas para centros de custo específicos e associe-os a Categorias de Custos da AWS para agrupar contas de unidades de negócios nos centros de custos de modo que o proprietário da unidade de negócios possa ver o consumo de várias contas em um só lugar.
- **Acesso:** rastreie as informações de faturamento de toda a organização no faturamento consolidado. Verifique se as partes interessadas e os proprietários de negócios certos têm acesso.
- **Controle:** crie mecanismos de governança eficazes com as barreiras de proteção certas para evitar cenários inesperados ao usar políticas de controle de serviços (SCPs), políticas de tags, políticas do IAM e alertas de orçamento. Por exemplo, é possível permitir que as equipes criem recursos específicos nas regiões preferidas somente usando mecanismos de controle eficazes, bem como evitar a criação de recursos sem uma tag específica (como centro de custos).
- **Estado atual:** configure um painel que mostra os níveis atuais de custo e uso. O painel deve estar disponível em um local altamente visível dentro do ambiente de trabalho, de modo semelhante a

um painel de operações. É possível exportar dados e usar o Painel de Custos e Uso do Hub de Otimização de Custos da AWS ou qualquer produto compatível para criar essa visibilidade. Talvez seja necessário criar painéis diferentes para pessoas diferentes. Por exemplo, o painel do gerente pode ser diferente de um painel de engenharia.

- **Notificações:** forneça notificações quando o custo ou o uso excederem os limites definidos e ocorrerem anomalias com o AWS Budgets ou a Detecção de Anomalias em Custos da AWS.
- **Relatórios:** resuma todas as informações de custos e uso. Aumente a conscientização e a responsabilização sobre seus gastos com a nuvem com dados de custos detalhados e alocáveis. Crie relatórios que sejam relevantes para a equipe que os consome e contenham recomendações.
- **Rastreamento:** mostre os custos e o uso atuais em relação a metas ou objetivos configurados.
- **Análise:** permita que os membros da equipe realizem análises personalizadas e profundas até a granularidade por hora, diária ou mensal com filtros diferentes (recurso, conta, tag etc.).
- **Inspeção:** mantenha-se em dia com suas oportunidades de otimização de custos e implantação de recursos. Receba notificações usando o Amazon CloudWatch, o Amazon SNS ou o Amazon SES para implantações de recursos na organização. Analise as recomendações de otimização de custos com o AWS Trusted Advisor ou o AWS Compute Optimizer.
- **Relatórios de tendências:** exiba a variabilidade de custo e uso ao longo do período necessário, com a granularidade necessária.
- **Previsões:** mostre os custos futuros estimados, faça uma estimativa do uso de recursos e gaste com painéis de previsão criados por você.

Você pode usar o [Hub de Otimização de Custos da AWS](#) para entender possíveis oportunidades de redução de custos consolidadas a partir de um local centralizado e criar exportações de dados para integração com o Amazon Athena. Você também pode usar o Hub de Otimização de Custos da AWS para implantar o Painel de Custos e Uso, que utiliza o Amazon QuickSight para análise interativa de custos e compartilhamento seguro de insights de custos.

Se você não conta com as habilidades essenciais ou a largura de banda necessária em sua organização, então poderá trabalhar com o [AWS ProServ](#), o [AWS Managed Services \(AMS\)](#) ou [parceiros da AWS](#). Você também pode usar ferramentas de terceiros, mas não se esqueça de validar a proposta de valor.

Etapas de implementação

- Permita o acesso baseado em equipe às ferramentas: configure suas contas e crie grupos que tenham acesso aos relatórios de custos e uso necessários para seus consumos e use o [AWS](#)

[Identity and Access Management](#) para [controlar o acesso](#) a ferramentas como o AWS Cost Explorer. Esses grupos devem incluir representantes de todas as equipes que possuem ou gerenciam uma aplicação. Isso garante que cada equipe tenha acesso às próprias informações de custo e uso para monitorar o consumo.

- Organize tags e categorias de custos: organize seus custos entre equipes, unidades de negócios, aplicações, ambientes e projetos. Use tags de recursos para organizar custos por tags de alocação de custos. Crie categorias de custos com base nas dimensões usando tags, contas, serviços etc. para mapear os custos.
- Configure o AWS Budgets: [configure o AWS Budgets](#) em todas as contas para suas workloads. Defina orçamentos para o gasto geral da conta e orçamentos para as workloads usando tags e categorias de custos. Configure notificações no AWS Budgets para receber alertas quando você exceder valores orçados ou quando os custos estimados excederem seus orçamentos.
- Configure a Detecção de Anomalias em Custos da AWS: use a [Detecção de Anomalias em Custos da AWS](#) para as contas, os serviços centrais ou as categorias de custos criadas para monitorar os custos e o uso e detectar gastos incomuns. É possível receber alertas individualmente em relatórios agregados, assim como alertas por e-mail ou em um tópico do Amazon SNS, o que permite analisar e determinar a causa-raiz de uma anomalia e identificar o fator que está aumentando o custo.
- Use ferramentas de análise custos: configure o [AWS Cost Explorer](#) para sua workload e contas para visualizar seus dados de custos para análise posterior. Crie um painel para a workload que rastreie o gasto geral, as principais métricas de uso da workload e a previsão de custos futuros com base nos seus dados de custo históricos.
- Use ferramentas de análise de redução de custos: use o Hub de Otimização de Custos da AWS para identificar oportunidades de economia com recomendações personalizadas, incluindo exclusão de recursos não utilizados, dimensionamento correto, Savings Plans, reservas e recomendações de otimizadores de computação.
- Configure ferramentas avançadas: opcionalmente, é possível criar recursos visuais para facilitar a análise interativa e o compartilhamento de informações sobre custos. Com as exportações de dados no Hub de Otimização de Custos da AWS, é possível criar um painel de custo e uso viabilizado pelo Amazon QuickSight para sua organização que fornece detalhes adicionais e granularidade. Você também pode implementar recursos avançados de análise usando exportações de dados no [Amazon Athena](#) para consultas avançadas e criar painéis no [Amazon QuickSight](#). Trabalhe com [Parceiros da AWS](#) para adotar soluções de gerenciamento de nuvem para monitoramento e otimização consolidados de faturas da nuvem.

Recursos

Documentos relacionados:

- [O que é o gerenciamento de custos da AWS Billing and Cost Management?](#)
- [Estabelecer seu ambiente de práticas recomendadas da AWS](#)
- [Práticas recomendadas para marcação de recursos da AWS com tags](#)
- [Marcar recursos da AWS](#)
- [AWS Cost Categories](#)
- [Analisar os custos com o AWS Budgets](#)
- [Analisar custos com o AWS Cost Explorer](#)
- [O que são exportações de dados da AWS?](#)

Vídeos relacionados:

- [Implantar Cloud Intelligence Dashboards](#)
- [Receber alertas sobre qualquer métrica ou KPI de FinOps ou otimização de custos](#)

Exemplos relacionados:

- [Painel de Custos e Uso baseado no Amazon QuickSight](#)
- [Workshop Governança de custos e uso na AWS](#)

COST03-BP06 Alocar custos com base nas métricas de workload

Aloque os custos da workload com base em métricas de uso ou resultados de negócios para medir o custo-benefício da workload. Implemente um processo para analisar os dados de custo e uso com serviços de análise, que podem fornecer informações e capacidade de estorno.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Otimizar os custos significa fornecer resultados de negócios com o menor preço, o que só pode ser alcançado por meio da alocação dos custos da workload com base nas métricas da workload (medidas pela eficiência da workload). Monitore as métricas de workload definidas por meio de

arquivos de log ou outro monitoramento de aplicações. Combine esses dados com os custos da workload, os quais podem ser obtidos examinando-se os custos com um valor de tag ou ID de conta específico. Execute essa análise por hora. Sua eficiência normalmente muda se você tem componentes de custo estático (por exemplo, um banco de dados de backend em execução de maneira permanente) com uma taxa de solicitações variável (por exemplo, picos de uso entre 9h e 17h e poucas solicitações à noite). Entender a relação entre os custos estáticos e variáveis ajuda você a concentrar suas atividades de otimização.

Criar métricas de workload para recursos compartilhados pode ser um desafio em comparação com recursos como aplicações em contêineres no Amazon Elastic Container Service (Amazon ECS) e no Amazon API Gateway. No entanto, existem algumas maneiras de categorizar o uso e rastrear os custos. Se precisar monitorar recursos compartilhados do AWS Batch e do Amazon ECS, você poderá habilitar os dados de alocação de custos divididos no AWS Cost Explorer. Com dados de alocação de custos divididos, você pode entender e otimizar o custo e o uso de suas aplicações em contêineres e alocar os custos das aplicações para entidades comerciais individuais com base na forma como os recursos compartilhados de computação e memória são consumidos.

Etapas de implementação

- Aloque custos a métricas da workload: usando as métricas definidas e a tags configuradas, crie uma métrica que combine a saída e o custo da workload. Use serviços de análise, como o Amazon Athena e o Amazon QuickSight, para criar um painel de eficiência para a workload geral e todos os componentes.

Recursos

Documentos relacionados:

- [Marcando recursos do AWS](#)
- [Analisar os custos com o AWS Budgets](#)
- [Analisar custos com Explorador de Custos](#)
- [Gerenciar Relatórios de Custos e Uso da AWS](#)

Exemplos relacionados:

- [Melhorar a visibilidade de custos do Amazon ECS e do AWS Batch com dados de alocação de custos divididos da AWS](#)

Desativar recursos

Depois de gerenciar uma lista de projetos, funcionários e recursos de tecnologia ao longo do tempo, você poderá identificar quais recursos não estão mais sendo usados e quais projetos não têm mais um proprietário.

Práticas recomendadas

- [COST04-BP01 Rastrear os recursos ao longo da vida útil](#)
- [COST04-BP02 Implementar um processo de desativação](#)
- [COST04-BP03 Desativar recursos](#)
- [COST04-BP04 Desativar recursos automaticamente](#)
- [COST04-BP05 Impor políticas de retenção de dados](#)

COST04-BP01 Rastrear os recursos ao longo da vida útil

Defina e implemente um método para rastrear recursos e suas associações com sistemas ao longo da vida útil. A marcação com tags pode ser usada para identificar a workload ou a função do recurso.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Desative recursos de workload que não são mais necessários. Um exemplo comum são os recursos usados para testes: após a conclusão do teste, os recursos podem ser removidos. O rastreamento de recursos com tags (e execução de relatórios sobre essas tags) pode ajudar a identificar ativos para desativação, pois eles não estarão em uso ou a licença deles expirará. Usar tags é uma maneira eficaz de rastrear recursos por meio da rotulagem do recurso com sua função ou uma data conhecida em que ele pode ser desativado. Os relatórios podem ser executados nessas tags. Os valores de exemplo para marcação de recursos são `feature-X testing` para identificar a finalidade do recurso em termos de ciclo de vida da workload. Outro exemplo é usar `LifeSpan` ou `TTL` para os recursos, como o nome e o valor da chave da tag a ser excluída, para definir o período ou o horário específico para a desativação.

Etapas de implementação

- Implemente um esquema de marcação com tags: implemente um esquema de marcação que identifique a workload à qual o recurso pertence, verificando se todos os recursos dentro da workload estão marcados da maneira apropriada. A marcação ajuda a categorizar os recursos

por finalidade, equipe, ambiente ou outros critérios relevantes para o seu negócio. Para obter mais detalhes sobre casos de uso, estratégias e técnicas de marcação, consulte [Práticas recomendadas de marcação com tags da AWS](#).

- Implemente o monitoramento de throughput ou saída da workload: implemente o monitoramento ou alarme de throughput da workload, iniciando nas solicitações de entrada ou na conclusão da saída. Configure-o para fornecer notificações quando saídas ou solicitações de workload caírem para zero, indicando que os recursos de workload não são mais usados. Incorpore um fator de tempo se a workload cair periodicamente para zero em condições normais. Para obter mais detalhes sobre recursos não utilizados ou subutilizados, consulte [Verificações de otimização de custos da AWS Trusted Advisor](#).
- Agrupe os recursos da AWS: crie grupos de recursos para seus recursos da AWS. Você pode usar o [AWS Resource Groups](#) para organizar e gerenciar seus recursos da AWS que fazem parte da mesma Região da AWS. É possível adicionar tags à maioria de seus recursos para ajudar a identificá-los e classificá-los em sua organização. Use o [Tag Editor](#) para adicionar tags em massa aos recursos compatíveis. Considere usar o [AWS Service Catalog](#) para criar, gerenciar e distribuir portfólios de produtos aprovados para usuários finais e gerenciar o ciclo de vida de seus produtos.

Recursos

Documentos relacionados:

- [AWS Auto Scaling](#)
- [AWS Trusted Advisor](#)
- [Verificações de otimização de custos da AWS Trusted Advisor](#)
- [Marcando recursos do AWS](#)
- [Publicar métricas personalizadas](#)

Vídeos relacionados:

- [Como otimizar custos usando o AWS Trusted Advisor](#)

Exemplos relacionados:

- [Organizar recursos da AWS](#)
- [Otimizar custos usando o AWS Trusted Advisor](#)

COST04-BP02 Implementar um processo de desativação

Implemente um processo para identificar e desativar recursos não utilizados.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Implemente um processo padronizado em toda a organização para identificar e remover recursos não utilizados. O processo deve definir a frequência das pesquisas e os processos para remover o recurso para verificar se todos os requisitos da organização foram atendidos.

Etapas de implementação

- Crie e implemente um processo de desativação: trabalhe com os proprietários e desenvolvedores da workload para criar um processo de desativação para a workload e seus respectivos recursos. O processo deve abranger o método para verificar se a workload está em uso e também se cada um dos recursos da workload está em uso. Detalhe as etapas necessárias para desativar o recurso, removendo-os do serviço e garantindo a conformidade com os requisitos normativos. Todos os recursos associados, como licenças ou armazenamento anexado, devem ser incluídos. Notifique os proprietários da workload de que o processo de desativação foi iniciado.

Use as seguintes etapas de desativação para obter orientações sobre o que deve ser verificado como parte do seu processo:

- Identifique os recursos a serem desativados: identifique os recursos que são elegíveis para desativação em sua Nuvem AWS. Registre todas as informações necessárias e agende a desativação. Em sua linha do tempo, certifique-se de considerar se (e quando) problemas inesperados surgirem durante o processo.
- Coordene e se comunique: trabalhe com os proprietários da workload para confirmar o recurso que será desativado
- Registre metadados e crie backups: registre metadados (como IPs públicos, região, AZ, VPC, sub-rede e grupos de segurança) e crie backups (como snapshots do Amazon Elastic Block Store ou AMI, exportação de chaves e exportação de certificados) se eles forem necessários para os recursos no ambiente de produção ou se forem recursos essenciais.
- Valide a infraestrutura como código: determine se os recursos foram implantados com o AWS CloudFormation, Terraform, AWS Cloud Development Kit (AWS CDK) ou qualquer outra ferramenta de implantação de infraestrutura como código para que possam ser reimplantados, se necessário.

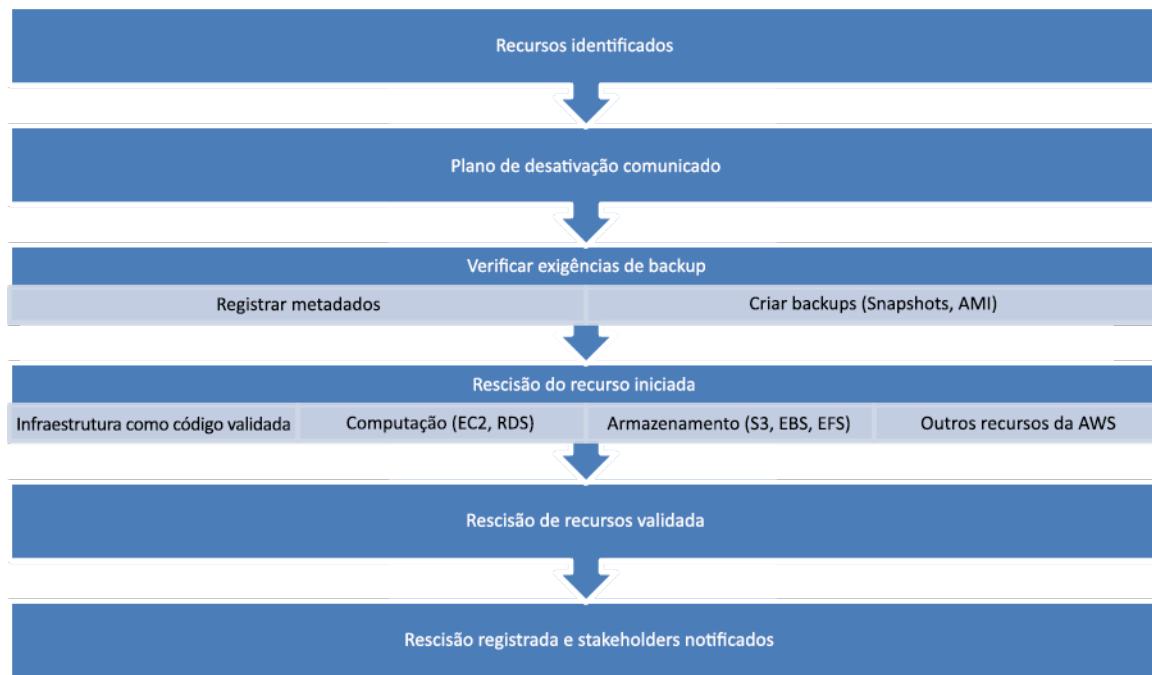
- **Impeça o acesso:** aplique controles restritivos por um período de tempo para evitar o uso de recursos enquanto você determina se o recurso é necessário. Verifique se o ambiente de recursos pode ser revertido para seu estado original, se necessário.
- **Siga seu processo interno de desativação:** siga as tarefas administrativas e o processo de desativação praticado por sua organização, como remover o recurso do domínio da organização, remover o registro DNS e remover o recurso das ferramentas de gerenciamento de configuração, de monitoramento, de automação e de segurança.

Se o recurso for uma instância do Amazon EC2, consulte a lista a seguir. [Para obter mais detalhes, consulte Como faço para excluir ou terminar meus recursos do Amazon EC2?](#)

- Pare ou encerre todas as suas instâncias e balanceadores de carga do Amazon EC2. Observação: as instâncias do EC2 permanecem visíveis no console por um breve período depois de terminadas. Você não será cobrado por instâncias que não estiverem em estado de execução
- Exclua sua infraestrutura do Auto Scaling.
- Libere todos os hosts dedicados.
- Exclua todos os volumes e snapshots do Amazon EBS.
- Libere todos os endereços IP elásticos.
- Cancele o registro das imagens de máquina da Amazon (AMIs).
- Encerre todos os ambientes do AWS Elastic Beanstalk.

Se o recurso for um objeto armazenado no e se você excluir um arquivo antes de atingir a duração mínima de armazenamento, uma taxa proporcional de exclusão antecipada será cobrada. A duração mínima de armazenamento do Amazon S3 Glacier depende da classe de armazenamento usada. Para obter um resumo da duração mínima de armazenamento para cada classe de armazenamento, consulte [Performance nas classes de armazenamento do Amazon S3](#). Para obter detalhes sobre como as taxas de exclusão antecipada são calculadas, consulte [Preços do Amazon S3](#).

O fluxograma simples do processo de desativação a seguir descreve as etapas de desativação. Antes de desativar recursos, verifique se os recursos que você identificou para desativação não estão sendo usados pela organização.



Fluxo de desativação de recursos.

Recursos

Documentos relacionados:

- [AWS Auto Scaling](#)
- [AWS Trusted Advisor](#)
- [AWS CloudTrail](#)

Vídeos relacionados:

- [Excluir a pilha do CloudFormation, mas reter alguns recursos](#)
- [Descobrir qual usuário iniciou a instância do Amazon EC2](#)

Exemplos relacionados:

- [Excluir ou encerrar recursos do Amazon EC2](#)
- [Descobrir qual usuário iniciou a instância do Amazon EC2](#)

COST04-BP03 Desativar recursos

Desative recursos iniciados por eventos, como auditorias periódicas ou alterações no uso. Em geral, a desativação pode ser realizada periodicamente e é manual ou automatizada.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A frequência e o esforço para pesquisar recursos não utilizados devem refletir as possíveis economias, portanto, uma conta com custo pequeno deve ser analisada com menos frequência do que uma conta com custos maiores. Pesquisas e eventos de desativação podem ser iniciados por alterações de estado na workload, como um produto que termina a vida útil ou é substituído. Pesquisas e eventos de desativação também podem ser iniciados por eventos externos, como alterações nas condições de mercado ou encerramento do produto.

Etapas de implementação

- Desative recursos: esse é o estágio de depreciação de recursos da AWS que não são mais necessários ou o término de um contrato de licenciamento. Conclua todas as verificações finais antes de avançar para o estágio de descarte e desativação de recursos para evitar interrupções indesejadas, como criar snapshots ou fazer backups. Usando o processo de desativação, desative cada um dos recursos que foram identificados como não utilizados.

Recursos

Documentos relacionados:

- [AWS Auto Scaling](#)
- [AWS Trusted Advisor](#)

Exemplos relacionados:

- [Laboratórios do Well-Architected: Desativar recursos \(Nível 100\)](#)

COST04-BP04 Desativar recursos automaticamente

Projete a workload para lidar normalmente com o encerramento de recursos ao identificar e desativar recursos não críticos, que não são necessários ou com baixa utilização.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Use a automação para reduzir ou remover os custos associados do processo de desativação. Projetar sua workload para executar a desativação automatizada reduzirá os custos gerais da workload durante sua vida útil. Você pode usar o [Amazon EC2 Auto Scaling](#) ou o [Application Auto Scaling](#) para realizar o processo de desativação. Você também pode implementar código personalizado usando a [API ou o SDK](#) para desativar recursos de workload automaticamente.

As [aplicações modernas](#) são criadas primeiro sem servidor, uma estratégia que prioriza a adoção de serviços sem servidor. A AWS desenvolveu [serviços com tecnologia sem servidor](#) para todas as três camadas da pilha: computação, integração e datastores. O uso da arquitetura sem servidor permitirá que você reduza os custos durante períodos de baixo tráfego com aumento e redução automáticos.

Etapas de implementação

- Implemente o Amazon EC2 Auto Scaling ou o Application Auto Scaling: para recursos compatíveis, configure-os com o Amazon EC2 Auto Scaling ou com o Application Auto Scaling. Esses serviços podem ajudar você a otimizar sua utilização e eficiência de custos ao consumir serviços da AWS. Quando a demanda cair, esses serviços removerão automaticamente qualquer excesso de capacidade de recursos para evitar gastos excessivos.
- Configure o CloudWatch para encerrar instâncias: as instâncias podem ser configuradas para encerrar usando os [alarmes do CloudWatch](#). Usando as métricas do processo de desativação, implemente um alarme com uma ação do Amazon Elastic Compute Cloud. Verifique a operação em um ambiente de não produção antes de implantar.
- Implemente código dentro da workload: você pode usar o AWS SDK ou a AWS CLI para desativar recursos da workload. Implemente código dentro da aplicação que se integre à AWS e encerre ou remova recursos que não são mais usados.
- Use serviços sem servidor: priorize a criação de [arquiteturas sem servidor](#) e [arquiteturas orientadas a eventos](#) na AWS para criar e suas aplicações. A AWS oferece vários serviços de tecnologia sem servidor que, inerentemente, fornecem automaticamente a utilização otimizada de recursos e a desativação automatizada (expansão e redução). Com aplicações sem servidor, a utilização de recursos é otimizada automaticamente e você nunca paga por provisionamento em excesso.

Recursos

Documentos relacionados:

- [Amazon EC2 Auto Scaling](#)
- [Conceitos básicos do Amazon EC2 Auto Scaling](#)
- [Application Auto Scaling](#)
- [AWS Trusted Advisor](#)
- [Tecnologia sem servidor na AWS](#)
- [Criar alarmes para parar, encerrar, reinicializar ou recuperar uma instância](#)
- [Adicionar ações de encerrar para a alarmes do Amazon CloudWatch](#)

Exemplos relacionados:

- [Agendar a exclusão automática de pilhas do AWS CloudFormation](#)
- [Laboratórios do Well-Architected: Desativar recursos automaticamente \(Nível 100\)](#)
- [AWS Auto Cleanup da Servian](#)

COST04-BP05 Impor políticas de retenção de dados

Defina as políticas de retenção de dados em recursos compatíveis para lidar com exclusão de objetos de acordo com os requisitos de suas organizações. Identifique e exclua recursos e objetos desnecessários ou órfãos que não sejam mais necessários.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Use políticas de retenção de dados e de ciclo de vida para reduzir os custos associados do processo de desativação e de armazenamento dos recursos identificados. A definição de suas políticas de retenção de dados e de ciclo de vida para realizar a exclusão e a migração automatizadas de classe de armazenamento reduzirá os custos gerais de armazenamento durante seu tempo de vida. Você pode usar o Amazon Data Lifecycle Manager para automatizar a criação e a exclusão de snapshots do Elastic Block Store e imagens de máquina (AMIs) baseadas no Amazon EBS, e usar o Amazon S3 Intelligent-Tiering ou uma configuração de ciclo de vida do Amazon S3 para gerenciar o ciclo de vida de seus objetos do Amazon S3. Também é possível implementar código personalizado usando a [API ou o SDK](#) para criar políticas de ciclo de vida e regras de política para que objetos sejam excluídos automaticamente.

Etapas de implementação

- Use o Amazon Data Lifecycle Manager: use políticas de ciclo de vida no Amazon Data Lifecycle Manager para automatizar a exclusão de snapshots do Amazon EBS e de AMIs baseadas no Amazon EBS.
- Defina a configuração do ciclo de vida em um bucket: use a configuração do ciclo de vida do Amazon S3 em um bucket para definir ações a serem tomadas pelo Amazon S3 durante o ciclo de vida de um objeto, bem como a exclusão no final do ciclo de vida do objeto, com base nos requisitos de sua empresa.

Recursos

Documentos relacionados:

- [AWS Trusted Advisor](#)
- [Amazon Data Lifecycle Manager](#)
- [Como definir uma configuração de ciclo de vida em um bucket do Amazon S3](#)

Vídeos relacionados:

- [Automatizar snapshots do Amazon EBS com o Amazon Data Lifecycle Manager](#)
- [Esvaziar um bucket do Amazon S3 usando uma regra de configuração de ciclo de vida](#)

Exemplos relacionados:

- [Esvaziar um bucket do Amazon S3 usando uma regra de configuração de ciclo de vida](#)
- [Laboratório do Well-Architected: Desativar recursos automaticamente \(Nível 100\)](#)

Recursos economicamente eficientes

Usar os serviços, os recursos e as configurações adequados para suas workloads é essencial para economias de custo. Considere o seguinte ao criar recursos economicamente eficientes:

Você pode usar arquitetos de soluções da AWS, soluções da AWS, arquiteturas de referência da AWS e parceiros da APN para obter ajuda para escolher uma arquitetura baseada no que você aprendeu.

Tópicos

- [Avaliar o custo ao selecionar serviços](#)
- [Selecionar o tipo de recurso, o tamanho e o número corretos](#)
- [Selecionar o melhor modelo de preços](#)
- [Planejar a transferência de dados](#)

Avaliar o custo ao selecionar serviços

Práticas recomendadas

- [COST05-BP01 Identificar os requisitos de custos da organização](#)
- [COST05-BP02 Analisar todos os componentes da workload](#)
- [COST05-BP03 Executar uma análise completa de cada componente](#)
- [COST05-BP04 Selecionar software com licenciamento econômico](#)
- [COST05-BP05 Selecionar os componentes desta workload para otimizar o custo alinhado com as prioridades da organização](#)
- [COST05-BP06 Realizar análises de custos para diferentes usos ao longo do tempo](#)

COST05-BP01 Identificar os requisitos de custos da organização

Trabalhe com os membros da equipe para definir o equilíbrio entre otimização de custos e outros pilares, como performance e confiabilidade, para essa workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Na maioria das organizações, o departamento de tecnologia da informação (TI) é composto de várias equipes pequenas, cada uma com sua própria agenda e área de foco, que refletem as especialidades e as habilidades dos respectivos membros. Você precisa compreender os objetivos, as prioridades e as metas gerais da organização e como cada departamento ou projeto contribui para esses objetivos. A categorização de todos os recursos essenciais, incluindo pessoal, equipamentos, tecnologia, materiais e serviços externos, é crucial para alcançar os objetivos organizacionais e um planejamento orçamentário abrangente. A adoção dessa abordagem sistemática para a identificação e a compreensão dos custos é fundamental para estabelecer um plano de custos realista e robusto para a organização.

ao selecionar serviços para a sua workload, é fundamental compreender as prioridades da sua organização. Crie um equilíbrio entre a otimização de custos e outros pilares do AWS Well-Architected Framework, como performance e confiabilidade. Esse processo deve ser conduzido de forma sistemática e regular para refletir as mudanças nos objetivos da organização, nas condições de mercado e na dinâmica operacional. Uma workload totalmente otimizada para custo é a solução mais alinhada aos requisitos da sua organização, mas não necessariamente o menor custo. Reúna-se com todas as equipes da organização, como produtos, negócios, técnicas e finanças, para coletar as informações. Avalie o impacto das compensações entre interesses concorrentes ou abordagens alternativas para ajudar a tomar decisões fundamentadas ao determinar onde concentrar as iniciativas ou escolher um plano de ação.

Por exemplo, a aceleração da velocidade de entrada no mercado de novos recursos pode ser enfatizada em relação à otimização de custos, ou você pode escolher um banco de dados relacional para dados não relacionais para simplificar o esforço de migração de um sistema, em vez de migrar para um banco de dados otimizado para seu tipo de dados e atualizar a aplicação.

Etapas de implementação

- Identifique os requisitos relacionados a custos da organização: reúna-se com membros da equipe da sua organização que incluam gerenciamento de produtos, proprietários de aplicações, equipes de desenvolvimento e operações, gerenciamento e finanças. Priorize os pilares do Well-Architected para essa workload e os respectivos componentes. O resultado deve ser uma lista ordenada dos pilares. Também é possível adicionar um peso a cada pilar para indicar quanto foco adicional ele tem ou o quanto semelhantes são os focos entre dois pilares.
- Resolva a dívida técnica e documente-a: durante a revisão da workload, resolva a dívida técnica. Documente um item de backlog para visitar a workload no futuro com o objetivo de refatorar ou

rearquitetar para otimizá-la ainda mais. É essencial comunicar claramente as compensações feitas para outras partes interessadas.

Recursos

Práticas recomendadas relacionadas:

- [REL11-BP07 Arquitetar o produto para cumprir as metas de disponibilidade e os acordos de serviço \(SLAs\) de tempo de atividade](#)
- [OPS01-BP06 Avaliar compensações](#)

Documentos relacionados:

- [Calculadora de custo total de propriedade \(TCO\) da AWS](#)
- [Classes de armazenamento do Amazon S3](#)
- [Produtos de nuvem](#)

COST05-BP02 Analisar todos os componentes da workload

Verifique se cada componente da workload é analisado, independentemente do tamanho ou dos custos atuais. O trabalho da análise deve refletir o benefício potencial, como os custos atuais e projetados.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Os componentes da workload, projetados para agregar valor comercial à organização, podem abranger vários serviços. Para cada componente, é possível escolher serviços específicos da Nuvem AWS para atender às necessidades dos negócios. Essa seleção pode ser influenciada por fatores como a familiaridade ou a experiência anterior com esses serviços.

Depois de identificar os requisitos de sua organização, conforme mencionado em [COST05-BP01 Identificar os requisitos de custos da organização](#), faça uma análise completa de todos os componentes de sua workload. Analise cada componente levando em conta os custos e os tamanhos atuais e projetados. Pense no custo da análise em relação a qualquer possível economia da workload ao longo do respectivo ciclo de vida. O trabalho despendido na análise de todos os componentes dessa workload deve corresponder às possíveis economias ou melhorias previstas da

otimização desse componente específico. Por exemplo, se o custo do recurso proposto for USD 10/mês e, sob as cargas previstas, não exceder USD 15/mês, gastar um dia de trabalho para reduzir os custos em 50% (USD 5 por mês) poderá exceder o benefício potencial durante a vida útil do sistema. Use uma estimativa baseada em dados mais rápida e eficiente para criar o melhor resultado geral para esse componente.

As workloads podem mudar ao longo do tempo, e o conjunto certo de serviços poderá não ser ideal se a arquitetura da workload ou o uso mudarem. A análise para seleção de serviços deve incorporar estados de workload e níveis de uso atuais e futuros. A implementação de um serviço para o estado ou o uso futuro da workload pode reduzir os custos gerais ao reduzir ou remover o esforço necessário para fazer alterações futuras. Por exemplo, usar o EMR sem servidor pode ser a escolha apropriada inicialmente. No entanto, à medida que o consumo desse serviço aumenta, a transição para o EMR no EC2 pode reduzir os custos desse componente da workload.

O [AWS Cost Explorer](#) e os AWS Cost and Usage Reports [CUR](#) podem analisar o custo de uma prova de conceito (PoC) ou um ambiente em execução. Você também pode usar o [AWS Pricing Calculator](#) para estimar os custos da workload.

Crie um fluxo de trabalho para ser seguido pelas equipes técnicas para analisar as workloads. Mantenha esse fluxo de trabalho simples, mas também abranja todas as etapas necessárias para garantir que as equipes entendam cada componente da workload e seus preços. Sua organização pode então acompanhar e personalizar esse fluxo de trabalho com base nas necessidades específicas de cada equipe.

1. Liste cada serviço em uso para sua workload: esse é um bom ponto de partida. Identifique todos os serviços em uso no momento e a origem dos custos.
2. Entenda como os preços funcionam para esses serviços: entenda o [modelo de preços](#) de cada serviço. Diferentes serviços da AWS têm modelos de preço diferentes com base em fatores como volume de uso, transferência de dados e preços específicos de recursos.
3. Concentre-se nos serviços que têm custos inesperados de workload e que não estão alinhados ao uso esperado e ao resultado comercial: identifique valores atípicos ou serviços em que o custo não seja proporcional ao valor ou ao uso utilizando o AWS Cost Explorer ou o AWS Cost and Usage Report. É importante correlacionar os custos com os resultados comerciais para priorizar os esforços de otimização.
4. AWS Cost Explorer, CloudWatch Logs, Logs de fluxo da VPC e Lente de Armazenamento do Amazon S3 para entender a causa-raiz desses altos custos: essas ferramentas são fundamentais no diagnóstico de custos elevados. Cada serviço oferece uma lente diferente para visualizar e analisar o uso e os custos. Por exemplo, o Explorador de Custos ajuda a determinar tendências

gerais de custos, o CloudWatch Logs fornece insights operacionais, os Logs de fluxo da VPC exibem o tráfego IP e a Lente de Armazenamento do Amazon S3 é útil para análises de armazenamento.

5. Use o AWS Budgets para definir orçamentos para determinados valores para serviços ou contas: definir orçamentos é uma forma proativa de gerenciar custos. Use o AWS Budgets para definir limites de orçamento personalizados e receber alertas quando os custos excederem esses limites.
6. Configure os alarmes do Amazon CloudWatch para enviar alertas de faturamento e uso: configure o monitoramento e alertas para métricas de custos e uso. Os alarmes do CloudWatch podem notificar você quando determinados limites forem violados, o que melhora o tempo de resposta da intervenção.

Promova melhorias notáveis e economias financeiras ao longo do tempo por meio da análise estratégica de todos os componentes da workload e independentemente de seus atributos atuais. O esforço investido nesse processo de análise deve ser deliberado, com consideração cuidadosa das vantagens que podem ser recebidas.

Etapas de implementação

- Liste os componentes da workload: crie uma lista dos componentes da sua workload. Use essa lista para verificar se cada componente foi analisado. O esforço despendido deve refletir a criticidade da workload conforme definido pelas prioridades da organização. Agrupe recursos de forma funcional para melhorar a eficiência (por exemplo, o armazenamento dos bancos de dados de produção, se houver vários bancos de dados).
- Priorize a lista de componentes: veja a lista de componentes e priorize-a em ordem de esforço. Normalmente, isso é feito por ordem de custos dos componentes, do mais caro para o mais barato, ou da criticidade, conforme definido pelas prioridades da organização.
- Faça a análise: para cada componente na lista, analise as opções e os serviços disponíveis e escolha a opção mais alinhada com suas prioridades organizacionais.

Recursos

Documentos relacionados:

- [AWS Pricing Calculator](#)
- [AWS Cost Explorer](#)
- [Classes de armazenamento do Amazon S3](#)

- [Produtos da Nuvem AWS](#)

Vídeos relacionados:

- [Série de otimização de custos da AWS: CloudWatch](#)

COST05-BP03 Executar uma análise completa de cada componente

Observe o custo geral de cada componente para a organização. Calcule o custo total de propriedade considerando o custo de operações e gerenciamento, especialmente ao usar serviços gerenciados pelo provedor de nuvem. O esforço de análise deve refletir o benefício potencial (por exemplo, o tempo gasto na análise é proporcional ao custo do componente).

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Considere a economia de tempo que permitirá que sua equipe se concentre na retirada de recursos de endividamento técnico, inovação, agregação de valor e criação de diferenciadores de negócios. Por exemplo, talvez você precise mover sem alterações (lift-and-shift) seu ambiente on-premises para a nuvem (também conhecido como redefinir a hospedagem) e otimizá-lo mais tarde. Vale a pena explorar as possíveis economias obtidas com o uso de serviços gerenciados na AWS que removem ou reduzem os custos de licença. Serviços gerenciados na AWS eliminam a sobrecarga operacional e administrativa da manutenção de um serviço, como aplicação de patches ou atualização do sistema operacional, e permitem que você se concentre na inovação e nos negócios.

Uma vez que os serviços gerenciados operam em escala da nuvem, eles podem oferecer menor custo por transação ou serviço. Você pode realizar possíveis otimizações para alcançar alguns benefícios tangíveis sem alterar a arquitetura principal da aplicação. Por exemplo, talvez você queira reduzir o tempo gasto gerenciando instâncias de banco de dados migrando para uma plataforma de banco de dados como serviço, como o [Amazon Relational Database Service \(Amazon RDS\)](#), ou migrando sua aplicação para uma plataforma totalmente gerenciada, como [AWS Elastic Beanstalk](#).

Geralmente, os serviços gerenciados têm atributos que podem ser definidos para garantir capacidade suficiente. Você deve definir e monitorar esses atributos para que sua capacidade em excesso seja mínima e a performance seja maximizada. Você pode modificar os atributos do AWS Managed Services usando o AWS Management Console ou as APIs e os SDKs da AWS para alinhar as necessidades de recursos à demanda em constante mudança. Por exemplo, é possível aumentar

ou diminuir o número de nós em um cluster do Amazon EMR (ou um cluster do Amazon Redshift) para aumentar ou reduzir a escala.

Você também pode unir várias instâncias em um recurso da AWS para ativar usos de maior densidade. Por exemplo, é possível provisionar vários bancos de dados pequenos em uma única instância de banco de dados do Amazon Relational Database Service (Amazon RDS). Conforme o uso aumenta, você pode migrar um dos bancos de dados para uma instância de banco de dados do Amazon RDS dedicada usando um processo de snapshot e restauração.

Ao provisionar workloads em serviços gerenciados, é necessário compreender os requisitos de ajuste da capacidade do serviço. Esses requisitos geralmente são tempo, esforço e qualquer impacto na operação normal da workload. O recurso provisionado deve permitir tempo para que as alterações ocorram. Provisione a sobrecarga necessária para permitir isso. O trabalho contínuo necessário para modificar os serviços pode ser reduzido a praticamente zero usando APIs e SDKs integrados a ferramentas de sistema e monitoramento como o Amazon CloudWatch.

Por exemplo, o [Amazon RDS](#), o [Amazon Redshift](#) e o [Amazon ElastiCache](#) fornecem um serviço de banco de dados gerenciado. O [Amazon Athena](#), o [Amazon EMR](#) e o [Amazon OpenSearch Service](#) oferecem um serviço de análise gerenciado.

O [AMS](#) é um serviço que opera a infraestrutura da AWS em nome de clientes e parceiros empresariais. Ele fornece um ambiente seguro e compatível no qual você pode implantar as workloads. O AMS usa modelos operacionais de nuvem empresarial com automação para permitir que você atenda aos requisitos da organização, migre para a nuvem mais rapidamente e reduza os custos de gerenciamento constantes.

Etapas de implementação

- Faça um análise rigorosa: usando a lista de componentes, trabalhe com cada componente da maior prioridade para a menor. Para componentes de prioridade maior e mais caros, execute análises adicionais e avalie todas as opções disponíveis e o impacto a longo prazo. Para componentes de prioridade menor, avalie se alterações no uso alterariam a prioridade do componente e, em seguida, execute uma análise do esforço apropriado.
- Compare recursos gerenciados e não gerenciados: considere o custo operacional dos recursos que você gerencia e compare-os com os recursos gerenciados da AWS. Por exemplo, analise seus bancos de dados em execução em instâncias do Amazon EC2 e compare-os com as opções do Amazon RDS (um serviço gerenciado pela AWS) ou do Amazon EMR em comparação com a execução do Apache Spark no Amazon EC2. Ao migrar de uma workload autogerenciada para uma workload totalmente gerenciada pela AWS, pesquise suas opções com cuidado. Os três

fatores mais importantes a serem considerados são o [tipo de serviço gerenciado](#) que você deseja usar, o processo que você usará para [migrar seus dados](#) e entender o [modelo de responsabilidade compartilhada da AWS](#).

Recursos

Documentos relacionados:

- [Calculadora de custo total de propriedade \(TCO\) da AWS](#)
- [Classes de armazenamento do Amazon S3](#)
- [Produtos da Nuvem AWS](#)
- [Modelo de responsabilidade compartilhada da AWS](#)

Vídeos relacionados:

- [Por que migrar para um banco de dados gerenciado?](#)
- [O que é o Amazon EMR e como posso usá-lo para processar dados?](#)

Exemplos relacionados:

- [Por que migrar para um banco de dados gerenciado?](#)
- [Consolide dados de bancos de dados SQL Server idênticos em um único banco de dados do Amazon RDS para SQL Server usando o AWS DMS](#)
- [Entregar dados em grande escala ao Amazon Managed Streaming for Apache Kafka \(Amazon MSK\)](#)
- [Migrar uma aplicação Web ASP.NET para o AWS Elastic Beanstalk](#)

COST05-BP04 Selecionar software com licenciamento econômico

Os softwares de código aberto eliminam os custos de licenciamento de software, o que pode contribuir com custos significativos para as workloads. Quando houver necessidade de um software licenciado, evite licenças vinculadas a atributos arbitrários, como CPUs, e procure aquelas que estejam vinculadas à saída ou aos resultados. O custo dessas licenças é mais próximo do benefício que elas oferecem.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

O código aberto originou-se no contexto do desenvolvimento de software para indicar que o software está em conformidade com determinados critérios de distribuição gratuita. O software de código aberto é composto de código-fonte que pode ser inspecionado, modificado e aprimorado por qualquer pessoa. Com base nos requisitos de negócios, nas habilidades dos engenheiros, no uso previsto ou em outras dependências tecnológicas, as organizações podem considerar o uso de software de código aberto na AWS para minimizar os custos de licença. Em outras palavras, o custo das licenças de software pode ser eliminado com o uso de [software de código aberto](#). Isso pode ter impacto significativo nos custos da workload à medida que seu tamanho é dimensionado.

Avalie os benefícios do software licenciado em relação ao custo total para otimizar a workload. Modele todas as alterações no licenciamento e como elas afetariam os custos da workload. Se um fornecedor alterar o custo da sua licença de banco de dados, investigue como isso afeta a eficiência geral da sua workload. Considere anúncios históricos de preços de seus fornecedores para identificar tendências de alterações de licenciamento em seus produtos. Os custos de licenciamento também podem ser dimensionados independentemente do throughput ou do uso, como licenças que escalam por hardware (licenças vinculadas à CPU). Essas licenças devem ser evitadas porque os custos podem aumentar rapidamente sem resultados correspondentes.

Por exemplo, operar uma instância do Amazon EC2 na região us-east-1 com um sistema operacional Linux permite reduzir os custos em aproximadamente 45% em comparação com a execução de outra instância do Amazon EC2 no Windows.

O [AWS Pricing Calculator](#) oferece uma maneira abrangente de comparar os custos de vários recursos com diferentes opções de licença, como instâncias do Amazon RDS e diferentes mecanismos de banco de dados. Além disso, o AWS Cost Explorer fornece uma perspectiva inestimável dos custos das workloads existentes, especialmente daquelas com licenças diferentes. Para gerenciamento de licenças, o [AWS License Manager](#) oferece um método simplificado para supervisionar e lidar com licenças de software. Os clientes podem implantar e operacionalizar o software de código aberto preferido na Nuvem AWS.

Etapas de implementação

- Analise as opções de licença: revise os termos de licenciamento do software disponível. Procure versões de código aberto que tenham a funcionalidade necessária e veja se os benefícios do software licenciado superam o custo. Termos favoráveis alinham o custo do software aos benefícios por ele oferecidos.

- Analise o provedor de software: revise todas as alterações históricas de preços ou licenciamento do fornecedor. Procure alterações que não estejam alinhadas aos resultados, como termos punitivos para execução em hardware ou plataformas de fornecedores específicos. Além disso, verifique como eles executam auditorias e as penalidades que poderiam ser impostas.

Recursos

Documentos relacionados:

- [Código aberto em AWS](#)
- [Calculadora de custo total de propriedade \(TCO\) da AWS](#)
- [Classes de armazenamento do Amazon S3](#)
- [Produtos de nuvem](#)

Exemplos relacionados:

- [Blogs de código aberto](#)
- [Blogs de código aberto da AWS](#)
- [Otimização e avaliação do licenciamento](#)

COST05-BP05 Selecionar os componentes desta workload para otimizar o custo alinhado com as prioridades da organização

Considere o custo ao selecionar todos os componentes para sua workload. Isso inclui o uso de serviços gerenciados e em nível de aplicação ou arquitetura sem servidor, contêineres ou orientada a eventos a fim de reduzir o custo geral. Minimize os custos de licença usando um software de código aberto ou que não tenha taxas de licença ou alternativas para reduzir os gastos.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Considere o custo de serviços e opções ao selecionar todos os componentes. Isso inclui o uso de serviços gerenciados e em nível de aplicação, como o [Amazon Relational Database Service](#) (Amazon RDS), [Amazon DynamoDB](#), [Amazon Simple Notification Service](#) (Amazon SNS) e [Amazon Simple Email Service](#) (Amazon SES) para reduzir o custo geral da organização.

Use contêineres e tecnologia sem servidor para computação, como o [AWS Lambda](#) e o [Amazon Simple Storage Service](#) (Amazon S3) para sites estáticos. Se possível, containerize sua aplicação e use serviços de contêiner gerenciados da AWS, como [Amazon Elastic Container Service](#) (Amazon ECS) ou [Amazon Elastic Kubernetes Service](#) (Amazon EKS).

Minimize os custos de licença usando software de código aberto ou software sem taxas de licença: por exemplo, Amazon Linux para workloads de computação ou migração de bancos de dados para o Amazon Aurora.

É possível usar serviços sem servidor ou em nível de aplicação, como [Lambda](#), [Amazon Simple Queue Service \(Amazon SQS\)](#), [Amazon SNS](#) e [Amazon SES](#). Esses serviços eliminam a necessidade de gerenciar um recurso e fornecem a função de execução de código, serviços de enfileiramento e entrega de mensagens. O outro benefício é que eles escalam a performance e o custo de acordo com o uso, permitindo a alocação e a atribuição eficientes de custos.

O uso da [arquitetura orientada a eventos](#) também é possível com serviços sem servidor. Arquiteturas orientadas a eventos são baseadas em push, então, tudo acontece sob demanda à medida que o evento se apresenta no roteador. Dessa forma, você não paga pela sondagem contínua para conferir um evento. Isso significa um consumo menor de largura de banda de rede, menor utilização de CPU, menor capacidade de frota ociosa e menos handshakes SSL/TLS.

Para obter mais informações sobre o Serverless, consulte o whitepaper [Aplicação sem servidor do Well-Architected](#).

Etapas de implementação

- Selecione cada serviço para otimizar o custo: usando sua análise e lista priorizada, selecione cada opção que fornece a melhor correspondência com suas prioridades organizacionais. Em vez de aumentar a capacidade para atender à demanda, considere outras opções que podem oferecer melhor performance por um custo menor. Por exemplo, se você precisar analisar o tráfego esperado para seus bancos de dados na AWS, considere aumentar o tamanho da instância ou usar serviços do Amazon ElastiCache (Redis ou Memcached) a fim de fornecer mecanismos em cache para seus bancos de dados.
- Avalie a arquitetura orientada a eventos: o uso de uma arquitetura sem servidor também permite criar uma arquitetura orientada a eventos para aplicações distribuídas e baseadas em microsserviço, o que ajuda a criar soluções escaláveis, resilientes, ágeis e econômicas.

Recursos

Documentos relacionados:

- [Calculadora de custo total de propriedade \(TCO\) da AWS](#)
- [AWS sem servidor](#)
- [O que é arquitetura orientada a eventos](#)
- [Classes de armazenamento do Amazon S3](#)
- [Produtos de nuvem](#)
- [Amazon ElastiCache \(Redis OSS\)](#)

Exemplos relacionados:

- [Conceitos básicos da arquitetura orientada a eventos](#)
- [Arquitetura orientada a eventos](#)
- [Como a Statsig funciona de forma 100x mais econômica usando o Amazon ElastiCache \(Redis OSS\)](#)
- [Práticas recomendadas para trabalhar com funções do AWS Lambda](#)

COST05-BP06 Realizar análises de custos para diferentes usos ao longo do tempo

As workloads podem mudar ao longo do tempo. Alguns serviços ou recursos são mais econômicos em diferentes níveis de uso. Ao executar a análise em cada componente ao longo do tempo e no uso projetado, a workload continua oferecendo um bom custo-benefício ao longo da vida útil.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

À medida que a AWS lança novos serviços e recursos, os serviços ideais para sua workload podem mudar. O esforço necessário deve refletir possíveis benefícios. A frequência da análise da workload depende dos requisitos da sua organização. Se for uma workload com custo significativo, implementar novos serviços mais cedo maximizará a redução de custos. Portanto, uma revisão mais frequente poderá ser vantajosa. Outro acionador da revisão é a alteração nos padrões de uso. Alterações significativas no uso podem indicar que serviços alternativos seriam opções melhores.

Se precisar mover dados para a Nuvem AWS, você poderá selecionar qualquer variedade de serviços oferecidos pela AWS e ferramentas de parceiros para ajudar a migrar seus conjuntos de dados, sejam eles arquivos, bancos de dados, imagens de máquina, volumes de bloco ou até backups de fita. Por exemplo, para mover um grande volume de dados para a AWS e dela ou processar dados na borda, você pode usar um dos dispositivos com propósito específico da AWS para mover petabytes de dados offline de forma econômica. Outro exemplo é relativo a taxas de transferência de dados mais altas, um serviço de conexão direta pode ser mais barato do que uma VPN, que fornece a conectividade consistente necessária para sua empresa.

Com base na análise de custos para uso diferente no decorrer do tempo, analise sua atividade de ajuste de escala. Analise o resultado para ver se a política de ajuste de escala pode ser ajustada para adicionar instâncias de vários tipos e opções de compra. Analise suas configurações para verificar se é possível reduzir o mínimo para atender às solicitações do usuário, mas com um tamanho de frota menor e adicionar mais recursos para atender à alta demanda esperada.

Realize análises de custos para diferentes usos ao longo do tempo, discutindo com as partes interessadas em sua organização e use o recurso de previsão do [AWS Cost Explorer](#) para prever o impacto potencial das mudanças no serviço. Monitore os acionadores de nível de uso utilizando o AWS Budgets, alarmes de faturamento do CloudWatch e o AWS Cost Anomaly Detection para identificar e implementar os serviços mais econômicos com maior rapidez.

Etapas de implementação

- Defina padrões de uso previstos: ao trabalhar com sua organização, como proprietários de produtos e marketing, documente quais serão os padrões de uso previstos e esperados para a workload. Converse com os stakeholders da empresa sobre aumentos de uso e custos históricos e previstos e garanta que os aumentos se alinhem com os requisitos da empresa. Identifique os dias, as semanas ou os meses em que você espera que mais usuários utilizem seus recursos da AWS, o que indica que você deve aumentar a capacidade dos recursos existentes ou adotar serviços adicionais a fim de reduzir o custo e aumentar a performance.
- Execute a análise de custos no uso previsto: usando os padrões de uso definidos, realize a análise em cada um desses pontos. O esforço de análise deve refletir o resultado provável. Por exemplo, se a alteração no uso for grande, uma análise completa deverá ser realizada para verificar quaisquer custos e alterações. Em outras palavras, quando o custo aumenta, o uso também deve aumentar para a empresa.

Recursos

Documentos relacionados:

- [Calculadora de custo total de propriedade \(TCO\) da AWS](#)
- [Classes de armazenamento do Amazon S3](#)
- [Produtos de nuvem](#)
- [Amazon EC2 Auto Scaling](#)
- [Migração de dados para nuvem](#)
- [AWS Snow Family](#)

Vídeos relacionados:

- [AWS OpsHub for Snow Family](#)

Selecionar o tipo de recurso, o tamanho e o número corretos

Ao selecionar os melhores tipo de recurso, tamanho e número de recursos, você atende aos requisitos técnicos com o recurso de menor custo. As atividades de dimensionamento correto levam em conta todos os recursos de uma workload, todos os atributos de cada recurso individual e o esforço envolvido na operação de dimensionamento correto. O dimensionamento correto pode ser um processo iterativo iniciado por alterações em padrões de uso e fatores externos, como quedas de preço da AWS ou novos tipos de recursos da AWS. O dimensionamento correto também poderá ser uma atividade única se o custo do esforço para dimensionar corretamente ultrapassar as economias potenciais durante a vida útil da workload.

Na AWS, há várias abordagens diferentes:

Práticas recomendadas

- [COST06-BP01 Realizar modelagem de custos](#)
- [COST06-BP02 Selecionar o tipo, o tamanho e o número do recurso com base nos dados](#)
- [COST06-BP03 Selecionar o tipo, tamanho e número do recurso automaticamente com base nas métricas](#)
- [COST06-BP04 Considerar o uso de recursos compartilhados](#)

COST06-BP01 Realizar modelagem de custos

Identifique os requisitos da organização (como as necessidades de negócios e os compromissos existentes) e realize a modelagem dos custos (custos gerais) da workload e de cada um de seus componentes. Realize atividades de referência para a workload sob diferentes cargas previstas e compare os custos. O esforço de modelagem deve refletir o benefício potencial. Por exemplo, o tempo gasto é proporcional ao custo do componente.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Execute a modelagem de custos para sua workload e cada um de seus componentes para entender o equilíbrio entre recursos e encontrar o tamanho correto para cada recurso na workload, considerando um nível específico de performance. O entendimento das considerações de custo pode embasar seu processo de tomada de decisão e caso de negócios organizacional ao avaliar os resultados da realização de valor para a implantação planejada da workload.

Realize atividades de referência para a workload sob diferentes cargas previstas e compare os custos. O esforço de modelagem deve refletir o benefício potencial. Por exemplo, o tempo gasto é proporcional ao custo do componente ou à economia prevista. Para obter as práticas recomendadas, consulte a seção [Revisão do pilar Eficiência de performance do AWS Well-Architected Framework](#).

Por exemplo, para criar modelagem de custos para uma workload que consiste em recursos computacionais, o [AWS Compute Optimizer](#) pode ajudar na modelagem de custos para workloads em execução. Ele fornece recomendações de dimensionamento correto para recursos de computação com base no uso histórico. Implante os CloudWatch Agents nas instâncias do Amazon EC2 para coletar métricas de memória que ajudam você com recomendações mais precisas no AWS Compute Optimizer. Essa é a fonte de dados ideal para recursos de computação, pois é um serviço gratuito e utiliza machine learning para fazer várias recomendações, dependendo dos níveis de risco.

Há [vários serviços](#) que você pode usar com logs personalizados como fontes de dados para operações de dimensionamento correto para outros serviços e componentes da workload, como [AWS Trusted Advisor](#), [Amazon CloudWatch](#) e [Amazon CloudWatch Logs](#). O AWS Trusted Advisor verifica os recursos e sinaliza aqueles com baixa utilização, o que pode ajudar você a dimensionar corretamente seus recursos e criar modelagem de custos.

Veja a seguir as recomendações para dados e métricas de modelagem de custo:

- O monitoramento deve refletir com precisão a experiência do usuário. Selecione a granularidade correta para o período e escolha com cuidado o máximo ou o 99º percentil, em vez da média.
- Selecione a granularidade correta para o período de análise necessário para cobrir todos os ciclos de workload. Por exemplo, se uma análise de duas semanas for realizada, talvez você esteja deixando passar um ciclo de alta utilização, o que pode levar a subprovisionamento.
- Escolha os serviços da AWS certos para sua workload planejada considerando seus compromissos existentes, modelos de preço selecionados para outras workloads e a capacidade de inovar com maior rapidez e concentrar-se em seu valor comercial principal.

Etapas de implementação

- Faça a modelagem de custos para recursos: implante a workload ou uma prova de conceito em uma conta separada com os tipos e tamanhos de recursos específicos a serem testados. Execute a workload com os dados de teste e registre os resultados de saída, bem como os dados de custo da hora em que o teste foi executado. Depois, reimplante a workload ou altere os tipos e tamanhos de recursos e execute novamente o teste. Inclua taxas de licença para todos os produtos que você pode usar com esses recursos e custos de operações estimados (mão de obra ou engenharia) para implantar e gerenciar esses recursos ao criar a modelagem de custo. Considere a modelagem de custo para um período (por hora, diária, anual ou três anos).

Recursos

Documentos relacionados:

- [AWS Auto Scaling](#)
- [Identificar oportunidades para dimensionar corretamente](#)
- [Recursos do Amazon CloudWatch](#)
- [Otimização de custos: dimensionamento correto do Amazon EC2](#)
- [AWS Compute Optimizer](#)
- [Calculadora de preços da AWS](#)

Exemplos relacionados:

- [Execute uma modelagem de custos baseada em dados](#)
- [Estime o custo das configurações de recursos da AWS planejados](#)

- [Escolher as ferramentas da AWS certas](#)

COST06-BP02 Selecionar o tipo, o tamanho e o número do recurso com base nos dados

Selecione o tamanho ou tipo do recurso com base nos dados sobre a workload e nas características do recurso. Por exemplo, computação, memória, throughput ou gravação intensiva. Essa seleção geralmente é feita usando uma versão anterior (on-premises) da workload, a documentação ou outras fontes de informações sobre a workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

O Amazon EC2 fornece uma ampla seleção de tipos de instância com diferentes níveis de capacidade de CPU, memória, armazenamento e rede para atender a diferentes casos de uso. Esses tipos de instância dispõem de diferentes combinações de capacidade de CPU, memória, armazenamento e rede, oferecendo versatilidade ao selecionar a combinação certa de recursos para os projetos. Eles são disponibilizados em vários tamanhos para que seja possível ajustar os recursos com base nas demandas da workload. Para determinar o tipo de instância necessário, reúna os detalhes dos requisitos do sistema da aplicação ou do software a ser executado na instância. Esses detalhes devem incluir:

- Sistema operacional
- Número de núcleos de CPU
- Núcleos de GPU
- Quantidade de memória do sistema (RAM)
- Tipo e espaço de armazenamento
- Requisito de largura de banda da rede

Identifique a finalidade dos requisitos de computação e a instância necessária e conheça as várias famílias de instâncias do Amazon EC2. A Amazon oferece as seguintes famílias de tipos de instância:

- Finalidade geral
- Otimizadas para computação

- Otimizadas para memória
- Otimizada para armazenamento
- Computação acelerada
- Otimizadas para HPC

Para uma compreensão mais profunda dos propósitos específicos e dos casos de uso que uma família de instâncias específica do Amazon EC2 pode atender, consulte [Tipos de instância da AWS](#).

A coleta dos requisitos do sistema é essencial para selecionar a família e o tipo de instância específicos que melhor atendem às suas necessidades. Os nomes dos tipos de instância são compostos do nome da família e do tamanho da instância. Por exemplo, a instância t2.micro é da família T2 e é de tamanho micro.

Selecione o tamanho ou o tipo de recurso com base na workload e nas características do recurso (por exemplo, computação, memória, throughput ou gravação intensiva). Essa seleção geralmente é feita usando a modelagem de custos, uma versão anterior da workload (como uma versão on-premises), a documentação ou outras fontes de informações sobre a workload (whitepapers ou soluções publicadas). O uso de calculadoras de preços ou de ferramentas de gerenciamento de custos da AWS pode ajudar a tomar decisões fundamentadas sobre tipos, tamanhos e configurações de instância.

Etapas de implementação

- Selecione recursos com base em dados: use seus dados de modelagem de custos para selecionar o nível de uso previsto da workload e escolha o tipo e o tamanho do recurso especificado. Com base nos dados da modelagem de custos, determine o número de CPUs virtuais, a memória total (GiB), o volume de armazenamento de instâncias local (GB), os volumes do Amazon EBS e o nível de performance da rede, levando em consideração a taxa de transferência de dados necessária para a instância. Sempre faça seleções com base em análise detalhada e em dados precisos para otimizar a performance e, ao mesmo tempo, gerenciar os custos de forma eficiente.

Recursos

Documentos relacionados:

- [AWS Tipos de instância](#)
- [AWS Auto Scaling](#)

- [Recursos do Amazon CloudWatch](#)
- [Otimização de custos: dimensionamento correto do EC2](#)

Vídeos relacionados:

- [Selecionar a instância certa do Amazon EC2 para suas workloads](#)
- [Dimensionar seus serviços da maneira certa](#)

Exemplos relacionados:

- [Agora é mais fácil descobrir e comparar os tipos de instância do Amazon EC2](#)

COST06-BP03 Selecionar o tipo, tamanho e número do recurso automaticamente com base nas métricas

Use métricas da workload em execução no momento para selecionar o tamanho e o tipo certos para otimizar o custo. Provisione adequadamente o throughput, o dimensionamento e o armazenamento para serviços de computação, armazenamento, dados e rede. Isso pode ser feito com um ciclo de comentários, como ajuste de escala automático ou por código personalizado na workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Crie um loop de comentários dentro da workload que usa métricas ativas da workload em execução para fazer alterações nessa workload. É possível usar um serviço gerenciado, por exemplo, o [AWS Auto Scaling](#), configurado por você para realizar as operações de dimensionamento certas. A AWS também fornece [APIs](#), [SDKs](#) e funcionalidades que permitem que os recursos sejam modificados com o mínimo esforço. Você pode programar uma workload para interromper e iniciar uma instância do Amazon EC2 para permitir uma alteração de tamanho ou tipo de instância. Isso fornece os benefícios do dimensionamento correto e, ao mesmo tempo, remove quase todo o custo operacional necessário para fazer a alteração.

Alguns serviços da AWS têm seleção automática de tipo ou tamanho, como o [Amazon Simple Storage Service Intelligent-Tiering](#). O Amazon S3 Intelligent-Tiering move automaticamente seus dados entre dois níveis de acesso: acesso frequente e acesso infrequente, com base em seus padrões de uso.

Etapas de implementação

- Aumente sua observabilidade configurando as métricas da workload: capture as principais métricas da workload. Essas métricas fornecem uma indicação da experiência do cliente, como a saída da workload, e se alinham às diferenças entre tipos e tamanhos de recursos, como uso de CPU e memória. Para recursos de computação, analise os dados de performance para dimensionar corretamente suas instâncias do Amazon EC2. Identifique instâncias ociosas e instâncias subutilizadas. As principais métricas a serem observadas são o uso da CPU e a utilização da memória (por exemplo, 40% de utilização da CPU em 90% do tempo, conforme explicado em [Dimensionamento correto com o AWS Compute Optimizer e utilização de memória habilitada](#)). Identifique instâncias com uso máximo de CPU e utilização de memória inferior a 40% em um período de quatro semanas. Essas são as instâncias que devem ser dimensionadas corretamente para reduzir os custos. Para recursos de armazenamento como o Amazon S3, você pode usar a [Lente de Armazenamento do Amazon S3](#), que permite ver 28 métricas em várias categorias no nível do bucket e 14 dias de dados históricos no painel por padrão. Você pode filtrar seu painel da Lente de Armazenamento do Amazon S3 por resumo e otimização de custos ou eventos para analisar métricas específicas.
- Veja as recomendações de dimensionamento correto: use as recomendações de dimensionamento correto no AWS Compute Optimizer e a ferramenta de dimensionamento correto do Amazon EC2 no console de gerenciamento de custos ou revise o dimensionamento correto dos recursos no AWS Trusted Advisor para fazer ajustes em sua workload. É importante usar as [ferramentas certas](#) ao dimensionar corretamente diferentes recursos e seguir as [diretrizes](#) de dimensionamento correto, seja uma instância do Amazon EC2, classes de armazenamento da AWS ou tipos de instância do Amazon RDS. Para recursos de armazenamento, é possível usar a Lente de Armazenamento do Amazon S3, que oferece visibilidade do uso de armazenamento de objetos e tendências de atividade, bem como faz recomendações acionáveis para otimizar custos e aplicar as práticas recomendadas de proteção de dados. Ao usar as recomendações contextuais que a [Lente de Armazenamento do Amazon S3](#) obtém da análise de métricas em toda a sua organização, você pode tomar medidas imediatas para otimizar seu armazenamento.
- Selecione o tipo e o tamanho do recurso automaticamente com base em métricas: usando as métricas de workload, selecione manual ou automaticamente os recursos da workload. Para recursos de computação, a configuração do AWS Auto Scaling ou a implementação de código dentro da aplicação pode reduzir o esforço necessário se alterações frequentes forem necessárias e, possivelmente, implementar alterações antes de um processo manual. Você pode iniciar e escalar automaticamente uma frota de instâncias sob demanda e instâncias spot em um único grupo do Auto Scaling. Além de receber descontos pelo uso de instâncias spot, você pode

usar instâncias reservadas ou um Savings Plan para receber taxas de desconto da definição de preço normal de instância sob demanda. Todos esses fatores combinados ajudam você a otimizar sua redução de custos para instâncias do Amazon EC2 e determinar a escala e a performance desejadas para a aplicação. Você também pode usar uma estratégia de [seleção de tipo de instância baseada em atributos \(ABS\)](#) em [grupos do Auto Scaling \(ASG\)](#), que permite expressar seus requisitos de instância como um conjunto de atributos, como vCPU, memória e armazenamento. É possível usar automaticamente os tipos de instância de geração mais recente quando eles são lançados e acessar uma variedade mais ampla de capacidade com instâncias spot do Amazon EC2. O Amazon EC2 Fleet e o Amazon EC2 Auto Scaling selecionam e executam instâncias que se ajustam aos atributos especificados, eliminando a necessidade de escolher manualmente os tipos de instância. Para recursos de armazenamento, você pode usar os recursos [Amazon S3 Intelligent Tiering](#) e [Amazon EFS Infrequent Access](#), que permitem selecionar automaticamente classes de armazenamento que proporcionam economia automática de custos de armazenamento quando os padrões de acesso aos dados mudam, sem impacto na performance ou sobrecarga operacional.

Recursos

Documentos relacionados:

- [AWS Auto Scaling](#)
- [Dimensionamento correto da AWS](#)
- [AWS Compute Optimizer](#)
- [Recursos do Amazon CloudWatch](#)
- [Configuração do CloudWatch](#)
- [Publicação de métricas personalizadas no CloudWatch](#)
- [Conceitos básicos do Amazon EC2 Auto Scaling](#)
- [Amazon S3 Storage Lens](#)
- [Amazon S3 Intelligent-Tiering](#)
- [Amazon EFS Infrequent Access](#)
- [Iniciar uma instância do Amazon EC2 usando o SDK](#)

Vídeos relacionados:

- [Dimensionar seus serviços da maneira certa](#)

Exemplos relacionados:

- [Seleção de tipo de instância baseada em atributos para Auto Scaling para Amazon EC2 Fleet](#)
- [Otimizar o Amazon Elastic Container Service para custos usando ajuste de escala agendado](#)
- [Ajuste de escala preditivo com o Amazon EC2 Auto Scaling](#)
- [Otimizar custos e obter visibilidade do uso com a Lente de Armazenamento do Amazon S3](#)
- [Laboratórios do Well-Architected: Recomendações de dimensionamento correto \(Nível 100\)](#)

COST06-BP04 Considerar o uso de recursos compartilhados

Para serviços já implantados no nível da organização para várias unidades de negócios, considere usar recursos compartilhados para aumentar a utilização e reduzir o custo total de propriedade (TCO). O uso de recursos compartilhados pode ser uma opção econômica para centralizar o gerenciamento e os custos ao usar soluções existentes, compartilhar componentes ou ambos. Gerencie funções comuns, como monitoramento, backups e conectividade, dentro dos limites de uma conta ou em uma conta dedicada. Também é possível reduzir os custos implementando padronização, reduzindo a ocorrência de duplicação e diminuindo a complexidade.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Nas situações em que várias workloads desempenham a mesma função, use soluções existentes e componentes compartilhados para melhorar o gerenciamento e otimizar os custos. Considere usar os recursos existentes (especialmente os compartilhados), como servidores de banco de dados de não produção ou serviços de diretórios, para reduzir os custos de nuvem, seguindo as práticas recomendadas de segurança e as regulamentações organizacionais. Para otimizar a obtenção de valor e a eficiência, é fundamental alocar os custos de volta (usando showback e chargeback) às áreas pertinentes da empresa que geram consumo.

Showback refere-se a relatórios que dividem os custos da nuvem em categorias atribuíveis, como consumidores, unidades de negócios, contas contábeis gerais ou outras entidades responsáveis. O objetivo do showback é mostrar para as equipes, unidades de negócios ou indivíduos os respectivos custos de consumo de recursos da nuvem.

Chargeback significa alocar os gastos com serviços centrais às unidades de custo com base em uma estratégia adequada para um processo específico de gerenciamento financeiro. Para os clientes,

o chargeback cobra o custo incorrido de uma conta de serviços compartilhados em diferentes categorias de custos financeiros adequadas para um processo de geração de relatórios de clientes. Ao estabelecer mecanismos de chargeback, você pode relatar os custos incorridos por diferentes unidades de negócios, produtos e equipes.

As workloads podem ser categorizadas como essenciais e não essenciais. Com base nessa classificação, use recursos compartilhados com configurações gerais para workloads menos essenciais. Para otimizar ainda mais os custos, reserve servidores dedicados exclusivamente para workloads essenciais. Compartilhe recursos ou provisione-os em várias contas para gerenciá-los de maneira eficiente. Mesmo em situações com ambientes distintos de desenvolvimento, teste e produção, o compartilhamento seguro é viável e não compromete a estrutura organizacional.

Para melhorar sua compreensão e otimizar os custos e o uso de aplicações em contêineres, utilize dados de alocação de custos divididos que ajudam a alocar os custos para entidades de negócios individuais com base na forma como a aplicação consome recursos compartilhados de computação e memória. Os dados de alocação de custos divididos ajudam você a obter showback e chargeback em nível de tarefa em workloads de contêiner executadas no Amazon Elastic Container Service (Amazon ECS) ou no Amazon Elastic Kubernetes Service (Amazon EKS).

Para arquiteturas distribuídas, crie uma VPC de serviços compartilhados que forneça acesso centralizado aos serviços compartilhados exigidos pelas workloads em cada uma das VPCs. Esses serviços compartilhados podem incluir recursos como serviços de diretório ou endpoints da VPC. Para reduzir as despesas administrativas e os custos, compartilhe os recursos de um local central em vez de criá-los em cada VPC.

Ao usar recursos compartilhados, é possível economizar nos custos operacionais, maximizar a utilização dos recursos e melhorar a consistência. Em um design de várias contas, é possível hospedar alguns serviços da AWS centralmente e acessá-los usando várias aplicações e contas em um hub para reduzir os custos. É possível usar o [AWS Resource Access Manager \(AWS RAM\)](#) para compartilhar outros recursos comuns, como [sub-redes de VPC e anexos do AWS Transit Gateway](#), [AWS Network Firewall](#) ou [pipelines do Amazon SageMaker AI](#). Em um ambiente de várias contas, use o AWS RAM para criar um recurso uma vez e compartilhá-lo com outras contas.

As organizações devem marcar os custos compartilhados de forma eficaz e verificar se não há uma parte significativa de seus custos sem marcação ou sem alocação. Se você não alocar os custos compartilhados de forma eficaz e ninguém assumir a responsabilidade pelo gerenciamento dos custos compartilhados, os custos de nuvem compartilhada podem sair do controle. Você deve saber onde custos foram incorridos nos níveis de recurso, workload, equipe ou organização, pois esse conhecimento aprimora sua compreensão do valor fornecido no nível aplicável quando comparado

aos resultados comerciais alcançados. Em última análise, as organizações se beneficiam de redução dos custos como resultado do compartilhamento da infraestrutura de nuvem. Incentive a alocação de custos em recursos de nuvem compartilhada para otimizar os gastos com a nuvem.

Etapas de implementação

- **Avalie os recursos existentes:** analise as workloads existentes que usam serviços semelhantes para sua workload. Dependendo dos componentes da workload, considere usar plataformas existentes se a lógica de negócios ou os requisitos técnicos permitirem.
- **Use o compartilhamento de recursos no AWS RAM e restrinja adequadamente:** use o AWS RAM para compartilhar recursos com outras contas da AWS em sua organização. Ao compartilhar recursos, você não precisa duplicar recursos em várias contas, o que minimiza a carga operacional da manutenção de recursos. Esse processo também ajuda você a compartilhar com segurança os recursos criados com perfis e usuários em sua conta e em outras Contas da AWS.
- **Marque recursos:** marque os recursos que são candidatos à geração de relatórios de custos e categorize-os dentro das categorias de custo. Ative essas tags de recursos relacionadas a custos para alocação de custos a fim de fornecer visibilidade do uso de recursos da AWS. Concentre-se em criar um nível adequado de granularidade com relação à visibilidade de custos e uso, além de incentivar comportamentos de consumo na nuvem por meio de relatórios de alocação de custos e rastreamento de KPIs.

Recursos

Práticas recomendadas relacionadas:

- [SEC03-BP08 Compartilhar recursos com segurança em sua organização](#)

Documentos relacionados:

- [O que é o AWS Resource Access Manager?](#)
- [Serviços da AWS que podem ser usados com o AWS Organizations](#)
- [Recursos compartilháveis da AWS](#)
- [Consultas do AWS Cost and Usage \(CUR\)](#)

Vídeos relacionados:

- [AWS Resource Access Manager: controle de acesso granular com permissões gerenciadas](#)

- [Como criar sua estratégia de alocação de custos da AWS](#)
- [Categorias de Custos da AWS](#)

Exemplos relacionados:

- [Chargeback de serviços compartilhados: um exemplo do AWS Transit Gateway](#)
- [Como criar um modelo de chargeback/showback para Savings Plans usando o CUR](#)
- [Usar o compartilhamento de VPC para uma arquitetura econômica de microsserviços de várias contas](#)
- [Melhorar a visibilidade de custos do Amazon EKS com dados de alocação de custos divididos da AWS](#)
- [Melhorar a visibilidade de custos do Amazon ECS e do AWS Batch com dados de alocação de custos divididos da AWS](#)

Selecionar o melhor modelo de preços

Executar a modelagem de custo da workload: considere os requisitos dos componentes da workload e entenda os possíveis modelos de preços. Defina o requisito de disponibilidade do componente. Determine se há vários recursos independentes que executam a função na workload e quais são os requisitos da workload ao longo do tempo. Compare o custo dos recursos usando o modelo de preços sob demanda padrão e outros modelos aplicáveis. Leve em consideração possíveis alterações nos recursos ou componentes da workload.

Executar análises regulares no nível da conta: a execução de uma modelagem de custos regular garante que as oportunidades de otimização em várias workloads possam ser implementadas. Por exemplo, se várias workloads usarem o modelo sob demanda, em um nível agregado, o risco de alteração será menor, e a implementação de um desconto baseado em compromisso atingirá um custo geral mais baixo. Recomenda-se realizar análises em ciclos regulares de duas semanas a um mês. Essa análise permite que você faça pequenas compras de ajuste para que a cobertura de seus modelos de preço continue a evoluir com suas workloads dinâmicas e os respectivos componentes.

Use a ferramenta de recomendações [AWS Cost Explorer](#) para encontrar oportunidades de descontos de compromisso.

Para encontrar oportunidades para workloads spot, use uma visualização por hora do uso geral e procure períodos regulares de uso ou elasticidade variáveis.

Modelos de preços: a AWS tem vários [modelos de preços](#) que permitem que você pague pelos seus recursos da maneira mais econômica que atenda às necessidades da sua organização. A seção a seguir descreve cada modelo de compra:

- Instâncias sob demanda
- Instâncias spot
- Descontos de compromisso: Savings Plans
- Descontos de compromisso: instâncias reservadas/capacidade
- Seleção geográfica
- Acordos e preços de terceiros

Sob demanda: esse é o modelo de preços padrão, com pagamento conforme o uso. Quando você usa recursos (por exemplo, instâncias do EC2 ou serviços como o DynamoDB sob demanda), você paga uma taxa fixa e não assume compromissos de longo prazo. É possível aumentar ou diminuir a capacidade de seus recursos ou serviços com base nas demandas da sua aplicação. O modelo Sob demanda tem uma taxa horária, mas, dependendo do serviço, pode ser cobrado em incrementos de 1 segundo (por exemplo, instâncias do Amazon RDS ou EC2 do Linux). O modelo Sob demanda é recomendado para aplicações com workloads de curto prazo (por exemplo, um projeto de quatro meses), com picos periódicos ou workloads imprevisíveis que não podem ser interrompidos. Ele também é adequado para workloads, como ambientes de pré-produção, que exigem tempos de execução ininterruptos, mas não são executados por tempo suficiente para um desconto de compromisso (Savings Plans ou instâncias reservadas).

Instâncias spot: uma [instância spot](#) é um recurso computacional extra do Amazon EC2 disponível com descontos de até 90% em preços sob demanda, sem necessidade de compromissos de longo prazo. Com as instâncias spot, você pode reduzir significativamente o custo de execução das aplicações ou escalar a capacidade computacional da aplicação para o mesmo orçamento. Ao contrário das instâncias sob demanda, as instâncias spot poderão ser interrompidas com um aviso de 2 minutos se o Amazon EC2 precisar da capacidade de volta ou se o preço da instância spot exceder o preço configurado. Em média, as instâncias spot são interrompidas em menos de 5% do tempo.

As instâncias spot são ideais quando há uma fila ou buffer implementado ou quando há vários recursos trabalhando de forma independente para processar as solicitações (por exemplo, processamento de dados do Hadoop). Normalmente, essas workloads são tolerantes a falhas, sem estado e flexíveis, como processamento em lotes, big data e análises, ambientes containerizados

e computação de alta performance (HPC). Workloads não críticas, como ambientes de teste e desenvolvimento, também são candidatas ao spot.

As instâncias spot também são integradas a vários serviços da AWS, como grupos do Amazon EC2 AutoScaling, Amazon EMR, Amazon Elastic Container Service (Amazon ECS) e AWS Batch.

Quando uma instância spot precisa ser recuperada, o Amazon EC2 envia um aviso de dois minutos por meio de um aviso de interrupção de instância spot entregue por meio do CloudWatch Events, bem como nos metadados da instância. Durante esse período de dois minutos, sua aplicação pode usar o tempo para salvar seu estado, drenar contêineres em execução, fazer upload de arquivos de log finais ou se remover de um balanceador de carga. Ao final dos dois minutos, você tem a opção de hibernar, interromper ou encerrar a instância spot.

Considere as seguintes práticas recomendadas ao adotar instâncias spot em suas workloads:

- Ser flexível no maior número possível de tipos de instância: seja flexível na família e no tamanho do tipo de instância para melhorar a probabilidade de atender aos requisitos de capacidade pretendidos, obter o menor custo possível e minimizar o impacto das interrupções.
- Ser flexível em relação a onde a workload será executada: a capacidade disponível pode variar de acordo com a zona de disponibilidade. Isso melhora a probabilidade de atender à capacidade pretendida ao tocar em vários grupos de capacidade sobressalente e fornece o menor custo possível.
- Projetar para continuidade: projete suas workloads para serem do tipo sem estado e tolerantes a falhas para que, se parte da capacidade do EC2 for interrompida, isso não afete a disponibilidade ou a performance da workload.
- Recomendamos o uso de instâncias spot em combinação com planos sob demanda e Savings Plans/instâncias reservadas para maximizar a otimização de custos da workload com a performance.

Descontos de compromisso - Savings Plans: a AWS fornece várias maneiras de reduzir seus custos reservando ou comprometendo-se a usar uma determinada quantidade de recursos e recebendo uma taxa com desconto para seus recursos. Um [Savings Plan](#) permite estabelecer um compromisso de gastos por hora por um ou três anos e receber preços com desconto em todos os seus recursos. Os Savings Plans oferecem descontos para serviços de computação da AWS, como Amazon EC2, AWS Fargate e AWS Lambda. Ao fazer o compromisso, você paga o valor desse compromisso a cada hora e ele é subtraído do uso sob demanda com a taxa de desconto. Por exemplo, você se compromete com USD 50 por hora e tem USD 150 por hora de uso sob demanda. Considerando

os preços dos Savings Plans, seu uso específico tem uma taxa de desconto de 50%. Portanto, seu compromisso de USD 50 cobre USD 100 de uso sob demanda. Você pagará USD 50 (compromisso) e USD 50 de uso sob demanda restante.

Os [Savings Plans para computação](#) são os mais flexíveis e oferecem um desconto de até 66%. Eles se aplicam automaticamente em zonas de disponibilidade, tamanho de instância, família de instâncias, sistema operacional, locação, região e serviço de computação.

[Savings Plans para instâncias](#) têm menos flexibilidade, mas fornecem uma taxa de desconto mais alta (até 72%). Eles se aplicam automaticamente em zonas de disponibilidade, tamanho de instância, sistema operacional e locação.

Existem três opções de pagamento:

- Sem pagamento adiantado: não há pagamento adiantado; você paga uma taxa horária reduzida a cada mês para o total de horas do mês.
- Pagamento adiantado parcial: fornece uma taxa de desconto mais alta do que Sem pagamento adiantado. Parte do uso é paga antecipadamente. Você então paga uma taxa horária reduzida menor a cada mês referente ao total de horas do mês.
- Pagamento adiantado integral: o uso de todo o período é pago antecipadamente, e nenhum outro custo é incorrido durante o restante do período de vigência pelo uso coberto pelo compromisso.

Você pode aplicar qualquer combinação dessas três opções de compra em suas workloads.

Os Savings Plans se aplicam primeiro ao uso na conta em que foram comprados, da porcentagem de desconto mais alta para a mais baixa e, em seguida, ao uso consolidado em todas as outras contas, da porcentagem de desconto mais alta para a mais baixa.

Recomenda-se comprar todos os Savings Plans em uma conta sem uso ou recursos, como a conta de gerenciamento. Isso garante que os Savings Plans apliquem as taxas de desconto mais altas em todo o seu uso, maximizando o valor do desconto.

As workloads e o uso normalmente mudam com o passar do tempo. Recomenda-se adquirir continuamente pequenas quantidades de compromissos de Savings Plans ao longo do tempo. Isso garante que você mantenha altos níveis de cobertura para maximizar seus descontos e que seus planos sempre atendam aos requisitos de workload e organização.

Não defina uma meta de cobertura em suas contas, devido à variação do desconto que é possível. A baixa cobertura não indica necessariamente um alto potencial de economia. A cobertura em

sua conta pode ser baixa, mas se seu uso for composto de instâncias pequenas, com um sistema operacional licenciado, a economia potencial poderá ser tão baixa quanto alguns %. Em vez disso, acompanhe e monitore as possíveis economias disponíveis na ferramenta de recomendação de Savings Plans. Analise com frequência as recomendações dos Savings Plans no Explorador de Custos (execute análises regulares) e continue a comprar compromissos até que as economias estimadas estejam abaixo do desconto necessário para a organização. Por exemplo, acompanhe e monitore se seus possíveis descontos permaneceram abaixo de 20%. Caso excedam esse valor, uma compra deverá ser feita.

Monitore a utilização e a cobertura, mas apenas para detectar alterações. Não aponte para uma porcentagem de utilização específica ou porcentagem de cobertura, pois isso não necessariamente escala com economia. Certifique-se de que uma compra de Savings Plans resulte em um aumento na cobertura e, se houver diminuição na cobertura ou utilização, garanta que eles sejam quantificados e conhecidos. Por exemplo, você migra um recurso de workload para um tipo de instância mais recente, o que reduz a utilização de um plano existente, mas o benefício de performance supera a redução de economia.

Descontos por compromisso – Instâncias reservadas/compromisso: de forma semelhante aos Savings Plans, as instâncias [reservadas](#) oferecem descontos de até 72% para um compromisso de executar uma quantidade mínima de recursos. As instâncias reservadas estão disponíveis para Amazon RDS, Amazon OpenSearch Service, Amazon ElastiCache, Amazon Redshift e DynamoDB. O Amazon CloudFront e o AWS Elemental MediaConvert também oferecem descontos quando você faz compromissos de uso mínimo. No momento, as instâncias reservadas estão disponíveis para o Amazon EC2. No entanto, os Savings Plans oferecem os mesmos níveis de desconto com maior flexibilidade e sem sobrecarga de gerenciamento.

As instâncias reservadas oferecem as mesmas opções de preços, sem adiantamento, pagamento adiantado parcial e pagamento adiantado integral, e os mesmos períodos de vigência de um ou três anos.

As instâncias reservadas podem ser adquiridas em uma região ou em uma zona de disponibilidade específica. Elas fornecem uma reserva de capacidade quando compradas em uma zona de disponibilidade.

O Amazon EC2 oferece RIs conversíveis, no entanto, os Savings Plans devem ser usados para todas as instâncias do EC2 devido à maior flexibilidade e redução dos custos operacionais.

O mesmo processo e métricas devem ser usados para rastrear e fazer compras de instâncias reservadas. Recomenda-se não rastrear a cobertura de RI em todas as suas contas. Também é

recomendável que a utilização percentual não seja monitorada ou rastreada. Em vez disso, visualize o relatório de utilização no Explorador de Custos e use a coluna de economia líquida na tabela. Se a economia líquida for um valor negativo significativamente grande, será necessário adotar medidas para corrigir a RI não utilizada.

Frota do EC2: uma [frota do EC2](#) é um recurso que permite definir uma capacidade computacional de destino e, em seguida, especificar os tipos de instância e o equilíbrio de instâncias sob demanda e spot para a frota. A frota do EC2 executará automaticamente a combinação de recursos de menor preço para atender à capacidade definida.

Seleção geográfica: quando você arquiteta suas soluções, uma prática recomendada é buscar colocar recursos computacionais mais perto dos usuários para fornecer menor latência e soberania de dados sólida. Para públicos globais, você deve usar vários locais para atender a essas necessidades. Você deve selecionar a localização geográfica que minimiza seus custos.

A infraestrutura global da Nuvem AWS é criada com base em [Regiões e zonas de disponibilidade](#). Uma região é um local físico do mundo onde há várias zonas de disponibilidade. As zonas de disponibilidade consistem em um ou mais data centers separados, cada um com energia, rede e conectividade redundantes, hospedados em instalações distintas.

Cada Região da AWS opera dentro das condições do mercado local, e os preços dos recursos são diferentes em cada região. Escolha uma região específica para operar um componente de sua solução completa para que você possa operar no menor preço possível globalmente. Você pode usar a Calculadora Mensal da AWS para estimar os custos da workload em várias regiões.

Acordos e definição de preço de terceiros: quando você utiliza soluções ou serviços de terceiros na nuvem, é importante que as estruturas de preços estejam alinhadas aos resultados da otimização de custos. Os preços deve ser dimensionados de acordo com os resultados e o valor fornecido. Um exemplo disso é um software que usa uma porcentagem das economias que ele fornece, quanto mais você economiza (resultado), mais ele cobra. Contratos que escalam com sua fatura normalmente não estão alinhados à otimização de custos, a menos que forneçam resultados para cada parte da sua fatura específica. Por exemplo, uma solução que fornece recomendações para o Amazon EC2 e cobra uma porcentagem de toda a sua fatura aumentará se você usar outros serviços para os quais ela não oferece nenhum benefício. Outro exemplo é um serviço gerenciado que é cobrado a uma porcentagem do custo dos recursos que são gerenciados. Um tamanho de instância maior pode não exigir necessariamente mais esforço de gerenciamento, mas será cobrado em um valor mais alto. Certifique-se de que essas disposições de preços de serviços incluam um programa de otimização de custos ou recursos em seu serviço para promover a eficiência.

Práticas recomendadas

- [COST07-BP01 Executar análise de modelo de preço](#)
- [COST07-BP02 Escolher regiões com base no custo](#)
- [COST07-BP03 Selecionar contratos de terceiros com termos econômicos](#)
- [COST07-BP04 Implementar modelos de preços para todos os componentes da workload](#)
- [COST07-BP05 Realizar análise de modelo de preços em nível da conta de gerenciamento](#)

COST07-BP01 Executar análise de modelo de preço

Analise cada componente da workload. Determine se o componente e os recursos serão executados por períodos estendidos (para descontos de compromisso) ou dinâmicos e curtos (para spot ou sob demanda). Execute uma análise da workload usando as recomendações nas ferramentas de gerenciamento de custos e aplique regras de negócios a essas recomendações para alcançar altos retornos.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

A AWS tem vários [modelos de preços](#) que permitem que você pague pelos seus recursos da maneira mais econômica que atenda às necessidades da sua organização. Trabalhe com suas equipes para determinar o modelo de preço mais apropriado. Com frequência, o modelo de preço consiste em uma combinação de várias opções, tal como determinado por seus requisitos de disponibilidade.

As instâncias sob demanda permitem que você pague pela capacidade computacional ou do banco de dados por hora ou segundo (mínimo de 60 segundos), dependendo das instâncias que você executa, sem compromisso de longo prazo ou pagamentos adiantados.

Os Savings Plans são um modelo de definição de preço flexível que oferece preços baixos no uso do Amazon EC2, Lambda e AWS Fargate em troca de um compromisso com uma quantidade constante de uso (medido em dólares por hora) por um período de vigência de 1 ou 3 anos.

As instâncias spot são um mecanismo de preços do Amazon EC2 que permite que você solicite capacidade computacional extra com desconto por hora (até 90% do preço sob demanda) sem compromisso prévio.

As instâncias reservadas oferecem até 75% de desconto mediante pagamento antecipado pela capacidade. Para obter mais detalhes, consulte [Otimização de custos com reservas](#).

Você pode optar por incluir um Savings Plan para os recursos associados aos ambientes de produção, qualidade e desenvolvimento. Como alternativa, como os recursos de sandbox só são ativados quando necessário, você pode escolher um modelo sob demanda para os recursos desse ambiente. Use [instâncias spot](#) da Amazon para reduzir os custos do Amazon EC2 ou use [Savings Plans para computação](#) para reduzir os custos do Amazon EC2, Fargate e Lambda. A ferramenta de recomendações do [AWS Cost Explorer](#) oferece oportunidades de descontos por compromisso com planos de poupança.

Se você já comprou [instâncias reservadas](#) para o Amazon EC2 no passado ou estabeleceu práticas de alocação de custos dentro da sua organização, poderá continuar usando as instâncias reservadas do Amazon EC2 por enquanto. Entretanto, recomendamos elaborar uma estratégia para usar Savings Plans no futuro como um mecanismo de redução de custos mais flexível. Você pode atualizar as recomendações de Savings Plans (SP) no AWS Cost Management para gerar novas recomendações de Savings Plans sempre que quiser. Use instâncias reservadas para reduzir os custos do Amazon RDS, Amazon Redshift, Amazon ElastiCache e Amazon OpenSearch Service. Os Savings Plans e as instâncias reservadas estão disponíveis em três opções: pagamento adiantado, pagamento adiantado parcial e sem pagamento adiantado. Use as recomendações de compra de IR e SP fornecidas no AWS Cost Explorer.

Para encontrar oportunidades para workloads spot, use uma visualização por hora do uso geral e procure períodos regulares de uso ou elasticidade variáveis. Você pode usar Instâncias Spot para várias aplicações flexíveis e tolerantes a falhas. Exemplos incluem servidores Web sem estado, endpoints de API, aplicações de big data e análise, workloads containerizadas, CI/CD e outras workloads flexíveis.

Analise suas instâncias do Amazon EC2 e do Amazon RDS para ver se elas podem ser desativadas quando não estiverem em uso (após o expediente e nos fins de semana). Essa abordagem permitirá que você reduza os custos em 70% ou mais em comparação a usá-las ininterruptamente. Se você tiver clusters do Amazon Redshift necessários apenas em momentos específicos, poderá pausar o cluster e, posteriormente, retomá-lo. Quando o cluster do Amazon Redshift ou a instância do Amazon EC2 e do Amazon RDS são interrompidos, o faturamento de computação é interrompido e somente se aplica a cobrança de armazenamento.

Observe que as [reservas de capacidade sob demanda](#) (ODCR) não são um desconto no preço. A reserva de capacidade é cobrada pela taxa sob demanda equivalente independentemente de você executar instâncias na capacidade reservada ou não. Elas devem ser consideradas quando você

precisa fornecer capacidade suficiente para os recursos que pretende executar. As ODCRs não precisam estar atreladas a compromissos de longo prazo, visto que elas podem ser canceladas quando não mais necessárias, mas elas também podem se beneficiar dos descontos que os Savings Plans ou as instâncias reservadas oferecem.

Etapas de implementação

- Analise a elasticidade da workload: usando a granularidade por hora no Explorador de Custos ou um painel personalizado, analise a elasticidade da workload. Procure alterações regulares no número de instâncias em execução. As instâncias de curta duração são candidatas a instâncias spot ou frota spot.
 - [Laboratório do Well-Architected: Explorador de Custos](#)
 - [Laboratório do Well-Architected: Visualização de custos](#)
- Revise os contratos de preços existentes: revise os contratos ou compromissos atuais quanto a necessidades de longo prazo. Analise o que você tem no momento e quanto esses compromissos estão em uso. Utilize descontos contratuais ou contratos empresariais preexistentes. Os [contratos empresariais](#) oferecem aos clientes a opção de personalizar os contratos que melhor atendem às suas necessidades. Com relação a compromissos de longo prazo, considere descontos de preço reservados, instâncias reservadas ou Savings Plans para o tipo específico de instância, a família de instâncias, a Região da AWS e as zonas de disponibilidade.
- Faça uma análise de desconto por compromisso: usando o Explorador de Custos em sua conta, revise as recomendações de Savings Plans e Reserved Instance. Para verificar se você implementou as recomendações corretas com os descontos e riscos necessários, siga os [laboratórios do Well-Architected](#).

Recursos

Documentos relacionados:

- [Como acessar as recomendações de instâncias reservadas](#)
- [Opções de compra de instâncias](#)
- [AWS Enterprise](#)

Vídeos relacionados:

- [Economizar até 90% e executar workloads de produção no spot](#)

Exemplos relacionados:

- [Laboratório do Well-Architected: Explorador de Custos](#)
- [Laboratório do Well-Architected: Visualização de custos](#)
- [Laboratório do Well-Architected: Modelos de preços](#)

COST07-BP02 Escolher regiões com base no custo

Os preços dos recursos podem ser diferentes em cada região. Identifique as diferenças de custo regionais e implante nas regiões com custos mais altos apenas se for necessário atender a requisitos de latência, residência e soberania de dados. A consideração do custo da região ajuda você a pagar o menor preço geral pela workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A [infraestrutura da Nuvem AWS](#) é global, hospedada em [vários locais em todo o mundo](#) e construída em torno de Regiões da AWS, zonas de disponibilidade, zonas locais, AWS Outposts e zonas do Wavelength. Uma região é um local físico no mundo, e cada região é uma área geográfica separada onde a AWS tem várias zonas de disponibilidade. As zonas de disponibilidade, que são locais isolados em cada região, consistem em um ou mais datacenters discretos, cada um com energia, rede e conectividade redundantes.

Cada Região da AWS opera de acordo com as condições do mercado local, e os preços dos recursos são diferentes em cada região devido às diferenças de custo de imóveis, fibra, eletricidade e impostos, por exemplo. Escolha uma região específica para operar um componente de sua solução completa para que você possa operar no menor preço possível globalmente. Use a [Calculadora de Preços da AWS](#) para calcular os custos de sua workload em várias regiões procurando serviços por tipo de local (região, zona do Wavelength e zona local) e região.

Ao projetar suas soluções, uma prática recomendada é buscar colocar os recursos de computação mais perto dos usuários para proporcionar menor latência e forte soberania de dados. Selecione a localização geográfica com base nos requisitos de segurança, performance, privacidade de dados e empresariais. Para aplicações com usuários finais globais, use vários locais.

Use regiões que ofereçam preços mais baixos por serviços da AWS para implantar suas workloads se não houver necessidade de atender a requisitos de privacidade de dados, segurança e

empresariais. Por exemplo, se sua região padrão for Ásia-Pacífico (Sydney) (ap-southwest-2) e não houver restrições (privacidade de dados, segurança, por exemplo) quanto ao uso de outras regiões, a implantação de instâncias não essenciais (desenvolvimento e teste) do Amazon EC2 na Leste dos EUA (Norte da Virgínia) (us-east-1) custará menos.

	<i>Conformidade</i>	<i>Latência</i>	<i>Custos</i>	<i>Serviços/recursos</i>
Região 1	✓	15 ms	\$\$	✓
Região 2	✓	20 ms	\$\$\$	X
Região 3	✓	80 ms	\$	✓
Região 4	✓	15 ms	\$\$	✓
Região 5	✓	20 ms	\$\$\$	X
Região 6	✓	15 ms	\$	✓
Região 7	✓	80 ms	\$	✓
Região 8	✓	15 ms	\$	X

Tabela de matriz de recursos de regiões

A tabela de matriz anterior mostra que a região 6 é a melhor opção para esse determinado cenário porque a latência é baixa em comparação a outras regiões, o serviço está disponível e é a região mais barata.

Etapas de implementação

- Revise os preços da Região da AWS: analise os custos da workload na região atual. Começando com os custos maiores por serviço e tipo de uso, calcule os custos nas outras regiões que estão disponíveis. Se a economia prevista ultrapassar o custo de mover o componente ou a workload, migre para a nova região.
- Revise os requisitos para implantações multirregionais: analise seus requisitos e obrigações empresariais (privacidade de dados, segurança ou performance) para descobrir se há restrições quanto ao uso de várias regiões. Se não houver obrigações que limitem o uso a uma única região, use várias regiões.

- Analise a transferência de dados necessária: considere os custos de transferência de dados ao selecionar regiões. Mantenha seus dados perto do seu cliente e dos recursos. Selecione Regiões da AWS mais baratas onde os dados fluam e haja transferência de dados mínima. Dependendo dos requisitos de sua empresa para transferência de dados, é possível usar o [Amazon CloudFront](#), [AWS PrivateLink](#), [AWS Direct Connect](#) e [AWS Virtual Private Network](#) para reduzir seus custos de rede, melhorar a performance e aprimorar a segurança.

Recursos

Documentos relacionados:

- [Como acessar as recomendações de instâncias reservadas](#)
- [Definição de preços do Amazon EC2](#)
- [Opções de compra de instâncias](#)
- [Tabelas de região](#)

Vídeos relacionados:

- [Economizar até 90% e executar workloads de produção no spot](#)

Exemplos relacionados:

- [Visão geral dos custos de transferência de dados para arquiteturas comuns](#)
- [Considerações de custo para implantações globais](#)
- [O que considerar ao selecionar uma região para suas workloads](#)
- [Laboratórios do Well-Architected: Uso restrito do serviço por região \(Nível 200\)](#)

COST07-BP03 Selecionar contratos de terceiros com termos econômicos

Contratos e termos econômicos garantem que o custo desses serviços seja dimensionado de acordo com os benefícios oferecidos. Selecione contratos e preços que possam ser escalados ao oferecerem benefícios adicionais à sua organização.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Há vários produtos no mercado que podem ajudar você a gerenciar os custos em ambientes de nuvem. Eles podem ter algumas diferenças em termos de recursos que dependem dos requisitos do cliente, como alguns que enfatizam a governança ou a visibilidade dos custos e outros a otimização de custos. Um fator muito importante para a eficácia da otimização e da governança de custos é usar a ferramenta certa com os recursos necessários e o modelo de preços correto. Esses produtos têm modelos de preços diferentes. Alguns aplicam determinada porcentagem de cobrança sobre sua fatura mensal, enquanto outros aplicam uma porcentagem sobre as economias obtidas. O ideal é pagar apenas pelo que você precisa.

Ao usar soluções ou serviços de terceiros na nuvem, é importante que as estruturas de preços estejam alinhadas aos resultados desejados. Os preços deve ser dimensionados de acordo com os resultados e o valor fornecido. Por exemplo, em software que leva uma porcentagem das economias que ele fornece, quanto mais você economiza (resultado), mais ele cobra. Os contratos de licença em que você paga mais conforme suas despesas aumentam nem sempre podem ser vantajosos em termos de otimização de custos. No entanto, se o fornecedor oferecer benefícios claros para todos os componentes da sua fatura, talvez essa taxa de ajuste de escala seja aceitável.

Por exemplo, uma solução que fornece recomendações para o Amazon EC2 e que aplica uma porcentagem de cobrança sobre toda a fatura poderá se tornar mais cara se você usar outros serviços que não oferecem nenhum benefício. Outro exemplo é um serviço gerenciado que é cobrado segundo uma porcentagem do custo dos recursos gerenciados. Um tamanho de instância maior pode não exigir necessariamente maior esforço de gerenciamento, mas pode custar mais caro. Verifique se essas disposições de preços de serviços incluem um programa ou recursos de otimização de custos no respectivo serviço para promover a eficiência.

Os clientes podem encontrar produtos mais avançados ou mais fáceis de usar no mercado. Você precisa considerar o custo desses produtos e avaliar possíveis resultados da otimização de custos a longo prazo.

Etapas de implementação

- **Analise contratos e termos de terceiros:** analise os preços nos contratos de terceiros. Execute modelagem para diferentes níveis de uso e leve em consideração novos custos, como o uso de novos serviços ou aumentos nos serviços atuais, devido ao crescimento da workload. Decida se os custos adicionais fornecem os benefícios necessários para a sua empresa.

Recursos

Documentos relacionados:

- [Como acessar as recomendações de instâncias reservadas](#)
- [Opções de compra de instâncias](#)

Vídeos relacionados:

- [Economizar até 90% e executar workloads de produção no spot](#)

COST07-BP04 Implementar modelos de preços para todos os componentes da workload

Os recursos em execução permanente devem utilizar capacidade reservada, como Savings Plans ou instâncias reservadas. A capacidade de curto prazo está configurada para usar instâncias spot ou frota spot. As instâncias sob demanda são usadas somente para workloads de curto prazo que não podem ser interrompidas e não são executadas por tempo suficiente para a capacidade reservada, entre 25% e 75% do período, dependendo do tipo do recurso.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Para melhorar o custo-benefício, a AWS fornece várias recomendações de compromisso com base no uso anterior. Essas recomendações podem ser usadas para compreender o que você pode economizar e como o compromisso será usado. É possível usar esses serviços como instâncias sob demanda ou spot ou assumir um compromisso por determinado período e reduzir os custos sob demanda com instâncias reservadas (RIs) e Savings Plans (SPs). É necessário compreender, além de cada componente da workload e dos vários serviços da AWS, os descontos de compromisso, as opções de compra e as instâncias spot desses serviços para otimizar a workload.

Considere os requisitos dos componentes da workload e informe-se sobre os diferentes modelos de preços desses serviços. Defina o requisito de disponibilidade desses componentes. Determine se há vários recursos independentes que executam a função na workload e quais são os requisitos da workload ao longo do tempo. Compare o custo dos recursos usando o modelo de preços sob demanda padrão e outros modelos aplicáveis. Leve em consideração possíveis alterações nos recursos ou componentes da workload.

Por exemplo, vamos analisar essa arquitetura de aplicações web na AWS. Esse exemplo de workload consiste em vários serviços da AWS, como Amazon Route 53, AWS WAF, Amazon CloudFront, instâncias do Amazon EC2, instâncias do Amazon RDS, balanceadores de carga, armazenamento do Amazon S3 e Amazon Elastic File System (Amazon EFS). É necessário analisar cada um desses serviços e identificar as possíveis oportunidades de redução de custos com diferentes modelos de preços. Alguns deles podem ser elegíveis para IRs ou SPs, e outros podem estar disponíveis apenas sob demanda. Conforme mostrado na imagem a seguir, alguns dos serviços da AWS podem ser comprometidos via IRs ou SPs.

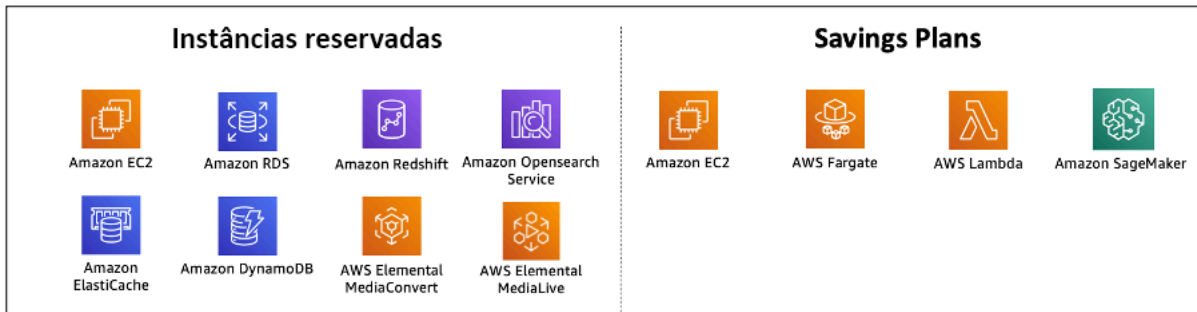


Tabela de serviços AWS da comprometidos via instâncias reservadas e Savings Plans

Etapas de implementação

- Implemente modelos de preços: usando seus resultados de análise, compre Savings Plans, instâncias reservadas ou implemente instâncias spot. Se esta for a sua primeira compra de compromisso, escolha as cinco ou dez principais recomendações da lista, monitore e analise os resultados de um ou dos dois próximos meses. O AWS Cost Management Console fornece orientações durante o processo. Analise as recomendações de IR ou de SP no console, personalize as recomendações (tipo, pagamento e prazo) e analise o compromisso por hora (por exemplo, USD 20 por hora) e adicione ao carrinho. Os descontos se aplicam automaticamente ao uso qualificado. Compre uma pequena quantidade de descontos de compromisso em ciclos regulares (por exemplo, a cada duas semanas ou mensalmente). Implemente instâncias spot para workloads que podem ser interrompidas ou que são sem estado. Finalmente, selecione instâncias sob demanda do Amazon EC2 e aloque recursos para os demais requisitos.
- Ciclo de revisão da workload: implemente um ciclo de análise da workload que analise especificamente a cobertura do modelo de preços. Assim que a workload tiver a cobertura necessária, compre descontos de compromisso adicionais parcialmente (a cada dois meses) ou conforme o uso da sua organização mudar.

Recursos

Documentos relacionados:

- [Entender as recomendações de Savings Plans](#)
- [Como acessar as recomendações de instâncias reservadas](#)
- [Como comprar instâncias reservadas](#)
- [Opções de compra de instâncias](#)
- [Instâncias spot](#)
- [Modelos de reserva para outros serviços da AWS](#)
- [Serviços com suporte de Savings Plans](#)

Vídeos relacionados:

- [Economizar até 90% e executar workloads de produção no spot](#)

Exemplos relacionados:

- [O que você deveria considerar antes de comprar Savings Plans?](#)
- [Como posso usar o Explorador de Custos para analisar meus gastos e uso?](#)

COST07-BP05 Realizar análise de modelo de preços em nível da conta de gerenciamento

Confira as ferramentas de gerenciamento de faturamento e de custos e veja os descontos recomendados com compromissos e reservas para realizar uma análise regular no nível da conta de gerenciamento.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Realizar modelagens de custos regularmente ajuda você a implementar oportunidades de otimização em várias workloads. Por exemplo, se várias workloads usarem instâncias sob demanda em um nível agregado, o risco de alteração será menor e a implementação de um desconto baseado em compromisso poderá atingir um custo geral mais baixo. Recomenda-se realizar análises em ciclos

regulares de duas semanas a um mês. Isso permite que você faça pequenas compras de ajuste para que a cobertura dos seus modelos de preços continue a evoluir com suas workloads dinâmicas e os respectivos componentes.

Use as ferramentas de recomendação da [AWS Cost Explorer](#) para encontrar oportunidades de descontos de compromisso em sua conta de gerenciamento. As recomendações em nível de conta de gerenciamento são calculadas considerando-se o uso em todas as contas da organização da AWS que têm instâncias reservadas (RI) ou Savings Plans (SP). Elas também são calculadas quando o compartilhamento de descontos é ativado para recomendar um compromisso que maximize a economia em todas as contas.

Embora a compra em nível da conta de gerenciamento seja otimizada para obter o máximo de economia em muitos casos, poderá haver situações em que você considera comprar SPs em nível da conta vinculada, como quando você deseja que os descontos se apliquem primeiro ao uso nessa conta vinculada específica. Conta-membro: as recomendações são calculadas no nível da conta individual ou conta-membro para maximizar a economia de cada conta. Se sua conta tiver compromissos de RI e SP, eles serão aplicados na seguinte ordem:

1. RI de zona
2. RI padrão
3. RI conversível
4. Savings Plans para instâncias
5. Savings Plans para computação

Se você comprar um SP em nível da conta de gerenciamento, a economia será aplicada com base na porcentagem de desconto mais alta para a mais baixa. Os SPs em nível da conta de gerenciamento examinam todas as contas vinculadas e aplicarão as economias sempre que o desconto for maior. Se desejar restringir onde as economias são aplicadas, você poderá comprar um Savings Plans em nível da conta vinculada e, sempre que a conta estiver executando serviços computacionais qualificados, o desconto será aplicado primeiro. Quando a conta não estiver executando serviços computacionais qualificados, o desconto será compartilhado entre as outras contas vinculadas na mesma conta de gerenciamento. O compartilhamento de descontos é ativado por padrão, mas pode ser desativado se necessário.

Em uma Família de Faturamento Consolidado, os Savings Plans são aplicados primeiro à conta do proprietário e depois a outras contas. Isso ocorre somente se você tiver o compartilhamento ativado. Seus Savings Plans são aplicados primeiro à maior porcentagem de economia. Se houver

vários usos com porcentagens iguais, os Savings Plans são aplicados ao primeiro uso com a menor taxa. Os Savings Plans continuam em vigor até que não haja mais usos restantes ou que seu compromisso seja esgotado. O uso restante é cobrado na tarifa sob demanda. Você pode atualizar as recomendações de Savings Plans no Gerenciamento de custos da AWS para gerar novas recomendações de Savings Plans sempre que quiser.

Depois de analisar a flexibilidade das instâncias, você pode confirmar de acordo com as recomendações. Crie uma modelagem de custos analisando os custos de curto prazo da workload com possíveis opções de recursos diferentes, analisando os modelos de preço da AWS e alinhando-os aos requisitos empresariais para encontrar o custo total de propriedade e oportunidades de [otimização de custos](#).

Etapas de implementação

Faça uma análise de desconto por compromisso: usando o Explorador de Custos em sua conta, revise as recomendações de Savings Plans e instâncias reservadas. Entenda as recomendações de Savings Plans e estime seus gastos mensais e as economias mensais. Examine as recomendações no nível da conta de gerenciamento, que são calculadas considerando o uso em todas as contas em sua organização da AWS que têm o compartilhamento de descontos de RI ou Savings Plans habilitado, com o intuito de obter o máximo de economia nas contas. É possível verificar se as recomendações corretas foram implementadas com os descontos e riscos necessários nos laboratórios do Well-Architected.

Recursos

Documentos relacionados:

- [Como a definição de preços da AWS funciona?](#)
- [Opções de compra de instâncias](#)
- [Visão geral dos Savings Plans](#)
- [Recomendações de Savings Plans](#)
- [Como acessar as recomendações de instâncias reservadas](#)
- [Conceitos básicos sobre a recomendação de Savings Plans](#)
- [Como os Savings Plans se aplicam ao seu uso do AWS](#)
- [Saving Plans com faturamento consolidado](#)
- [Ativar descontos compartilhados de instâncias reservadas e Savings Plans](#)

Vídeos relacionados:

- [Economizar até 90% e executar workloads de produção no spot](#)

Exemplos relacionados:

- [Laboratório do AWS Well-Architected: Modelos de preços \(Nível 200\)](#)
- [Laboratórios do AWS Well-Architected: Análise de modelo de preços \(Nível 200\)](#)
- [O que devo considerar antes de comprar um Savings Plans?](#)
- [Como posso usar Savings Plans sucessivos para reduzir o risco do compromisso?](#)
- [Quando usar instâncias spot](#)

Planejar a transferência de dados

Uma vantagem da nuvem é que ela é um serviço de rede gerenciado. Não há mais a necessidade de gerenciar e operar uma frota de switches, roteadores e outros equipamentos de rede associados. Os recursos de rede na nuvem são consumidos e pagos da mesma forma que você paga por CPU e por armazenamento: apenas pelo que usa. O uso eficiente de recursos de rede é necessário para otimização de custos na nuvem.

Práticas recomendadas

- [COST08-BP01 Executar a modelagem de transferência de dados](#)
- [COST08-BP02 Selecionar os componentes para otimizar o custo da transferência de dados](#)
- [COST08-BP03 Implementar serviços para reduzir custos de transferência de dados](#)

COST08-BP01 Executar a modelagem de transferência de dados

Reúna os requisitos da organização e execute a modelagem de transferência de dados da workload e de cada um dos componentes. Isso identifica o menor ponto de custo para os requisitos atuais de transferência de dados.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Ao projetar uma solução na nuvem, as taxas de transferência de dados geralmente são negligenciadas devido ao hábito de projetar a arquitetura usando datacenters on-premises ou à falta de conhecimento. As taxas de transferência de dados na AWS são determinadas pela origem, pelo destino e pelo volume do tráfego. A consideração dessas taxas durante a fase de projeto pode resultar em redução de custos. É muito importante compreender onde ocorre a transferência de dados na workload, o custo da transferência e os respectivos benefícios associados para estimar com precisão o custo total de propriedade (TCO). Isso permite que você tome uma decisão embasada para modificar ou aceitar a decisão de arquitetura. Por exemplo, você pode ter uma configuração de várias zonas de disponibilidade na qual replica dados entre as zonas de disponibilidade.

Você modela os componentes dos serviços que transferem os dados na workload e conclui que esse é um custo aceitável (de modo semelhante ao pagamento por computação e armazenamento nas duas zonas de disponibilidade) para alcançar a confiabilidade e a resiliência necessárias. Modele os custos em diferentes níveis de uso. O uso da workload pode mudar ao longo do tempo, e diferentes serviços podem ser mais econômicos em diferentes níveis.

Ao modelar a transferência de dados, considere a quantidade de dados ingeridos e a origem desses dados. Além disso, considere a quantidade de dados processados e a quantidade de armazenamento ou capacidade computacional necessária. Durante a modelagem, siga as práticas recomendadas de rede para sua arquitetura de workload a fim de otimizar os possíveis custos de transferência de dados.

O AWS Pricing Calculator pode ajudar você a ver os custos estimados de serviços específicos da AWS e da transferência de dados esperada. Se você já tiver uma workload em execução (para fins de teste ou em um ambiente de pré-produção), use o [AWS Cost Explorer](#) ou o [AWS Cost and Usage Report](#) (CUR) para entender e modelar seus custos de transferência de dados. Configure uma prova de conceito (PoC) ou teste sua workload e execute um teste com uma carga simulada realista. É possível modelar seus custos em diferentes demandas de workload.

Etapas de implementação

- Identifique os requisitos: qual é a meta principal e os requisitos de negócios para a transferência planejada de dados entre a origem e o destino? Quais são os resultados comerciais esperados no final? Reúna os requisitos de negócios e defina o resultado esperado.
- Identifique a origem e o destino: quais são a fonte e o destino dos dados para a transferência de dados, como dentro de Regiões da AWS, para serviços da AWS ou para fora da Internet?

- [Transferência de dados dentro de uma Região da AWS](#)
- [Transferência de dados entre Regiões da AWS](#)
- [Transferência de dados para a internet](#)
- Identifique as classificações de dados: qual é a classificação de dados para essa transferência de dados? De que tipo são esses dados? Qual é o tamanho dos dados? Com que frequência os dados devem ser transferidos? Os dados são sigilosos?
- Identifique serviços ou ferramentas da AWS a serem usados: que serviços da AWS são usados para essa transferência de dados? É possível usar um serviço já provisionado para outra workload?
- Calcule os custos de transferência de dados: use o modelo de transferência de dados de [Preços da AWS](#) que você criou anteriormente para calcular os custos de transferência de dados para a workload. Calcule os custos da transferência de dados em diferentes níveis de uso, tanto para aumentos quanto para reduções no uso da workload. Quando houver várias opções para a arquitetura da workload, calcule o custo de cada uma delas a título de comparação.
- Vincule os custos aos resultados: para cada custo de transferência de dados incorrido, especifique o resultado que ele atinge para a workload. Se a transferência for entre componentes, poderá ser para desacoplamento; se for entre zonas de disponibilidade, poderá ser para redundância.
- Crie uma modelagem de transferência de dados: depois de reunir todas as informações, crie uma modelagem de transferência de dados de base conceitual para vários casos de uso e workloads diferentes.

Recursos

Documentos relacionados:

- [Soluções de armazenamento em cache da AWS](#)
- [Definição de preço do AWS](#)
- [Definição de preços do Amazon EC2](#)
- [Preços da Amazon VPC](#)
- [Conceitos básicos das taxas de transferência de dados](#)

Vídeos relacionados:

- [Monitorar e otimizar os custos da transferência de dados](#)

- [Aceleração de transferências do S3](#)

Exemplos relacionados:

- [Visão geral dos custos de transferência de dados para arquiteturas comuns](#)
- [Recomendações da AWS para redes](#)

COST08-BP02 Selecionar os componentes para otimizar o custo da transferência de dados

Todos os componentes são selecionados, e a arquitetura é projetada para reduzir os custos de transferência de dados. Isso inclui o uso de componentes como otimização de rede de longa distância (WAN) e configurações de várias zonas de disponibilidade (AZ).

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A arquitetura da transferência de dados minimiza os custos da transferência de dados. Isso pode envolver o uso de redes de entrega de conteúdo para localizar os dados mais perto dos usuários ou o uso de links de rede dedicados do ambiente on-premises para a AWS. Você também pode usar a otimização de WAN e a otimização de aplicações para reduzir a quantidade de dados transferidos entre componentes.

Ao transferir dados para a Nuvem AWS ou dentro dela, é essencial conhecer o destino com base em diversos casos de uso, a natureza dos dados e os recursos de rede disponíveis para selecionar os serviços certos da AWS e otimizar a transferência de dados. A AWS oferece uma variedade de serviços de transferência de dados personalizados para diversos requisitos de migração de dados. Selecione as opções corretas de [armazenamento de dados](#) e [transferência de dados](#) com base nas necessidades comerciais da sua organização.

Ao planejar ou analisar a arquitetura da workload, considere o seguinte:

- Use endpoints da VPC dentro da AWS: um endpoint da VPC permite conexões privadas entre a VPC e os serviços da AWS compatíveis. Isso permite evitar o uso da internet pública, o que pode resultar em custos de transferência de dados.
- Use um gateway NAT: use um [gateway NAT](#) para que as instâncias em uma sub-rede privada possam se conectar a serviços fora da VPC. Verifique se os recursos por trás do gateway NAT que

enviam mais tráfego estão na mesma zona de disponibilidade do gateway NAT. Caso contrário, crie novos gateways NAT na mesma zona de disponibilidade do recurso para reduzir as taxas de transferência de dados entre AZs.

- O uso de AWS Direct Connect ignora a Internet pública e estabelece uma conexão direta e privada entre sua rede on-premises e a AWS. Isso pode ser mais econômico e consistente do que transferir grandes volumes de dados pela internet.
- Evite transferir dados entre fronteiras regionais: as transferências de dados entre Regiões da AWS (de uma região para outra) normalmente incorrem em cobranças. A decisão de seguir um caminho multirregional deve ser muito cuidadosa. Para obter mais detalhes, consulte [Cenários multirregionais](#).
- Monitore a transferência de dados: use o Amazon CloudWatch e os [Logs de fluxo da VPC](#) para capturar detalhes sobre sua transferência de dados e uso da rede. Analise as informações de tráfego de rede capturadas nas VPCs, como o endereço IP ou o intervalo de entrada e saída das interfaces de rede.
- Analise o uso da rede: use ferramentas de medição e geração de relatórios AWS Cost Explorer, como CUDOS Dashboards ou CloudWatch, para entender o custo de transferência de dados da sua workload.

Etapas de implementação

- Selecione os componentes para a transferência de dados: usando a modelagem de transferência de dados explicada em [COST08-BP01 Executar a modelagem de transferência de dados](#), concentre-se em onde estão os maiores custos de transferência de dados ou onde eles estariam se o uso da workload mudasse. Procure arquiteturas alternativas ou componentes adicionais que removam ou reduzam a necessidade da transferência de dados (ou que diminuam o custo).

Recursos

Práticas recomendadas relacionadas:

- [COST08-BP01 Executar a modelagem de transferência de dados](#)
- [COST08-BP03 Implementar serviços para reduzir custos de transferência de dados](#)

Documentos relacionados:

- [Migração de dados para nuvem](#)

- [Soluções de armazenamento em cache da AWS](#)
- [Entregar conteúdo com mais rapidez com o Amazon CloudFront](#)

Exemplos relacionados:

- [Visão geral dos custos de transferência de dados para arquiteturas comuns](#)
- [Dicas de otimização de rede da AWS](#)
- [Otimizar a performance e reduzir os custos de análise de rede com os Logs de fluxo da VPC no formato Apache Parquet](#)

COST08-BP03 Implementar serviços para reduzir custos de transferência de dados

Implemente serviços para reduzir os custos da transferência de dados. Por exemplo, é possível usar locais da borda ou redes de entrega de conteúdo (CDN) para fornecer conteúdo aos usuários finais, criar camadas de cache na frente de servidores de aplicações ou bancos de dados e usar conexões de rede dedicadas em vez de VPN para conectividade com a nuvem.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Existem vários serviços da AWS que podem ajudar a otimizar o uso da transferência de dados pela rede. Dependendo da arquitetura da nuvem e dos componentes e tipo da workload, esses serviços podem ajudar na compactação, no armazenamento em cache e no compartilhamento e na distribuição do tráfego na nuvem.

- O [Amazon CloudFront](#) é uma rede de entrega de conteúdo global que entrega dados com baixa latência e altas velocidades de transferência. Ele armazena dados em cache em pontos de presença no mundo inteiro, o que reduz a carga sobre seus recursos. Ao usar o CloudFront, você pode reduzir o trabalho administrativo para entregar conteúdo a grandes números de usuários globalmente com latência mínima. O [pacote promocional de segurança](#) pode ajudar você a economizar até 30% do uso do CloudFront se você planeja aumentar o uso ao longo do tempo.
- O [AWS Direct Connect](#) permite a você estabelecer uma conexão de rede dedicada com a AWS. Isso pode reduzir os custos de rede, aumentar a largura de banda e fornecer uma experiência de rede mais consistente do que conexões baseadas na Internet.

- O [AWS VPN](#) permite estabelecer uma conexão segura e privada entre a rede privada e a rede global da AWS. Ele é ideal para pequenos escritórios ou parceiros de negócios porque oferece conectividade simplificada, além de ser um serviço totalmente gerenciado e elástico.
- Os [endpoints da VPC](#) permitem a conectividade entre os serviços da AWS sobre redes privadas e podem ser usados para reduzir os custos de transferência de dados pública e [gateways NAT](#). Os [endpoints da VPC do gateway](#) não têm cobranças por hora e oferecem suporte ao Amazon S3 e ao Amazon DynamoDB. Os [endpoints da VPC de interface](#) são fornecidos pelo [AWS PrivateLink](#) e têm uma taxa horária e custo de uso por GB.
- Os [gateways NAT](#) fornecem ajuste de escala e gerenciamento integrados, reduzindo os custos, em comparação com uma instância NAT independente. Coloque os gateways NAT nas mesmas zonas de disponibilidade das instâncias de alto tráfego e pense no uso de endpoints da VPC para as instâncias que precisam acessar o Amazon DynamoDB ou o Amazon S3 a fim de reduzir os custos de transferência e processamento de dados.
- Use dispositivos [AWS Snow Family](#) que tenham recursos de computação para coletar e processar dados na borda. Os dispositivos AWS Snow Family ([Snowcone](#), [Snowball](#) e [Snowmobile](#)) permitem que você mova petabytes de dados para a Nuvem AWS de forma econômica e offline.

Etapas de implementação

- Implemente os serviços: selecione os serviços de rede aplicáveis da AWS com base no serviço e no tipo de workload usando a modelagem de transferência de dados e revisando os logs de fluxo da VPC. Veja onde estão os maiores custos e os maiores fluxos de volume. Revise os serviços da AWS e avalie se algum deles reduz ou remove a transferência, especificamente a entrega de conteúdo e as redes. Procure também serviços de armazenamento em cache em que há acesso repetido aos dados ou grandes quantidades de dados.

Recursos

Documentos relacionados:

- [AWS Direct Connect](#)
- [Explorar os produtos da AWS](#)
- [Soluções de armazenamento em cache da AWS](#)
- [Amazon CloudFront](#)
- [AWS Snow Family](#)

- [Pacote Promocional de Segurança do Amazon CloudFront](#)

Vídeos relacionados:

- [Monitorar e otimizar os custos da transferência de dados](#)
- [Série de otimização de custos da AWS: CloudFront](#)
- [Como posso reduzir os custos da transferência de dados para meu gateway NAT?](#)

Exemplos relacionados:

- [Chargeback de serviços compartilhados: um exemplo do AWS Transit Gateway](#)
- [Como entender os detalhes da transferência de dados da AWS com base no relatório de custos e uso utilizando consultas do Athena e o QuickSight](#)
- [Visão geral dos custos de transferência de dados para arquiteturas comuns](#)
- [Usar o AWS Cost Explorer para analisar custos de transferência de dados](#)
- [Otimizar os custos das arquiteturas da AWS utilizando os recursos do Amazon CloudFront](#)
- [Como posso reduzir os custos da transferência de dados para meu gateway NAT?](#)

Gerenciar recursos de demanda e fornecimento

Ao migrar para a nuvem, você paga apenas pelo que precisa. Você pode fornecer recursos para atender à demanda da workload no momento que eles são necessários, eliminando a necessidade de provisionamento em excesso dispendioso e com desperdício de recursos. Você também pode modificar a demanda usando um controle de utilização, um buffer ou uma fila para suavizar a demanda e atendê-la com menos recursos.

Os benefícios econômicos da oferta just-in-time devem ser equilibrados em relação à necessidade de provisionar para compensar falhas de recursos, alta disponibilidade e tempo de provisionamento. Dependendo de sua demanda (fixa ou variável), planeje criar métricas e automação que garantam que o gerenciamento de seu ambiente seja mínimo, mesmo conforme você ajusta a escala. Ao modificar a demanda, você deve saber o atraso aceitável e máximo que a workload pode permitir.

Na AWS, é possível usar várias abordagens diferentes para gerenciar a demanda e fornecer recursos. As seções a seguir descrevem como usar essas abordagens.

Práticas recomendadas

- [COST09-BP01 Realizar uma análise sobre a demanda da workload](#)
- [COST09-BP02 Implementar um buffer ou controle de utilização para gerenciar a demanda](#)
- [COST09-BP03 Fornecer recursos dinamicamente](#)

COST09-BP01 Realizar uma análise sobre a demanda da workload

Analise a demanda da workload ao longo do tempo. Garanta que a análise cubra tendências sazonais e represente com precisão as condições operacionais durante toda a vida útil da workload. O trabalho de análise deve refletir o benefício potencial (por exemplo, se o tempo gasto é proporcional ao custo da workload).

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Analisar a demanda de workload para computação em nuvem envolve entender os padrões e as características das tarefas de computação que são iniciadas no ambiente de nuvem. Essa análise ajuda os usuários a otimizar a alocação de recursos, gerenciar custos e verificar se a performance atende aos níveis exigidos.

Conhecer os requisitos da workload. Os requisitos da organização devem indicar os tempos de resposta da workload para solicitações. O tempo de resposta pode ser usado para determinar se a demanda é gerenciada ou se a oferta de recursos deve ser alterada para atender à demanda.

A análise deve incluir a previsibilidade e a repetibilidade da demanda, a taxa de alteração na demanda e a quantidade de alteração na demanda. Realize a análise durante um período longo o suficiente para incorporar qualquer variação sazonal, como processamento de fim de mês ou picos de fim de ano.

O trabalho de análise deve refletir os possíveis benefícios da implementação do ajuste de escala. Observe o custo total esperado do componente e os aumentos ou diminuições no uso e no custo durante a vida útil da workload.

Veja abaixo alguns aspectos importantes a serem considerados ao realizar a análise da demanda de workload para computação em nuvem:

1. Métricas de utilização de recursos e performance: analise como os recursos da AWS estão sendo usados ao longo do tempo. Determine padrões de uso de pico e fora do pico para otimizar as estratégias de alocação e ajuste de escala de recursos. Monitore métricas de performance, como tempos de resposta, latência, throughput e taxas de erro. Essas métricas ajudam a avaliar a integridade geral e a eficiência da infraestrutura de nuvem.
2. Comportamento de ajuste de escala de usuários e aplicações: entenda o comportamento do usuário e como ele afeta a demanda da workload. Examinar os padrões de tráfego de usuários ajuda a aprimorar a entrega de conteúdo e a capacidade de resposta das aplicações. Analise como as workloads escalam com o aumento da demanda. Determine se os parâmetros de ajuste de escala automático estão configurados de forma correta e eficaz para lidar com flutuações de carga.
3. Tipos de workload: identifique os diferentes tipos de workload em execução na nuvem, como processamento em lote, processamento de dados em tempo real, aplicação web, bancos de dados ou machine learning. Cada tipo de workload pode ter requisitos de recursos e perfis de performance diferentes.
4. Acordos de serviço (SLA): compare a performance real com os SLAs para garantir a conformidade e identificar áreas que precisam ser aprimoradas.

Você pode usar o [Amazon CloudWatch](#) para coletar e rastrear métricas, coletar e monitorar arquivos de log, definir alarmes e reagir automaticamente a alterações nos seus recursos da AWS. O Amazon

CloudWatch pode ser usado para fornecer visibilidade sobre a utilização de recursos, a performance de aplicações e o status operacional em todo o sistema.

Com o [AWS Trusted Advisor](#), é possível provisionar os recursos seguindo as práticas recomendadas para melhorar a performance e a confiabilidade do sistema, aumentar a segurança e procurar oportunidades de economia. Também é possível desativar o uso e as instâncias de não produção e usar o Amazon CloudWatch e o Auto Scaling para equiparar aumentos ou reduções na demanda.

Finalmente, você pode usar o [AWS Cost Explorer](#) ou o [Amazon QuickSight](#) com o arquivo do AWS Cost and Usage Report (CUR) ou os logs da aplicação para realizar análises avançadas da demanda de workload.

No geral, uma análise abrangente da demanda da workload permite que as organizações tomem decisões embasadas sobre provisionamento, ajuste de escala e otimização de recursos, o que melhora a performance, o custo-benefício e a satisfação do usuário.

Etapas de implementação

- **Analise dados da workload existente:** analise dados da workload existente, das versões anteriores da workload ou dos padrões de uso previstos. Use o Amazon CloudWatch, arquivos de log e dados de monitoramento para obter informações sobre como a workload foi usada. Analise um ciclo completo da workload e colete dados para alterações sazonais, como eventos de fim de mês ou de ano. O esforço refletido na análise deve refletir as características da workload. Deve-se concentrar o maior esforço em workloads de alto valor com as maiores alterações na demanda. Por outro lado, deve-se concentrar o menor esforço em workloads de baixo valor que tenham alterações mínimas na demanda.
- **Preveja a influência externa:** encontre membros da equipe de toda a organização que possam influenciar ou alterar a demanda na workload. Equipes comuns seriam de vendas, marketing ou desenvolvimento de negócios. Trabalhe com elas para saber os ciclos com os quais operam e se há eventos que possam alterar a demanda da workload. Preveja a demanda da workload com esses dados.

Recursos

Documentos relacionados:

- [Amazon CloudWatch](#)
- [AWS Trusted Advisor](#)

- [AWS X-Ray](#)
- [AWS Auto Scaling](#)
- [Agendador de instâncias da AWS](#)
- [Conceitos básicos do Amazon SQS](#)
- [AWS Cost Explorer](#)
- [Amazon QuickSight](#)

Vídeos relacionados:

Exemplos relacionados:

- [Monitorar, rastrear e analisar para fins de otimização de custos](#)
- [Pesquisar e analisar logs no CloudWatch](#)

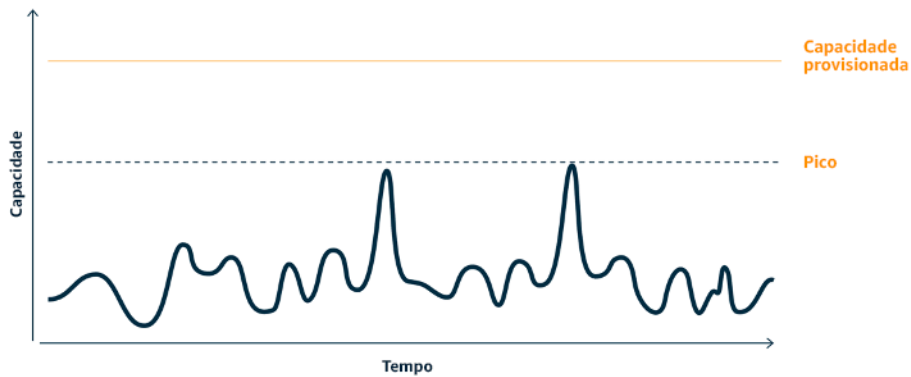
COST09-BP02 Implementar um buffer ou controle de utilização para gerenciar a demanda

O armazenamento em buffer e o controle de utilização modificam a demanda na workload, suavizando todos os picos. Implemente o controle de utilização quando seus clientes realizarem novas tentativas. Implemente o armazenamento em buffer para armazenar a solicitação e adiar o processamento até um momento posterior. Verifique se os controles de utilização e buffers estão projetados para que os clientes recebam uma resposta no tempo necessário.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A implementação de um buffer ou controle de utilização é crucial na computação em nuvem para gerenciar a demanda e reduzir a capacidade provisionada necessária para a workload. Para obter a performance ideal, é essencial avaliar a demanda total, incluindo os picos, o ritmo das mudanças nas solicitações e o tempo de resposta necessário. Quando os clientes têm a capacidade de reenviar solicitações, é prático aplicar o controle de utilização. Entretanto, para clientes que não têm funcionalidades de repetição, a abordagem ideal é implementar uma solução de buffer. Esses buffers agilizam o influxo de solicitações e otimizam a interação de aplicações com velocidades operacionais variadas.



Curva de demanda com dois picos distintos que exigem alta capacidade provisionada.

Considere uma workload com a curva de demanda mostrada na figura anterior. Essa workload tem dois picos e, para lidar com eles, é provisionada a capacidade de recurso mostrada pela linha laranja. Os recursos e a energia usados para essa workload não são indicados pela área abaixo da curva da demanda, mas pela área abaixo da linha da capacidade provisionada, visto que é preciso ter capacidade provisionada para lidar com esses dois picos. Nivelar a curva da demanda pode ajudar você a reduzir a capacidade provisionada para uma workload e a diminuir o respectivo impacto ambiental. Para suavizar o pico, considere implementar uma solução de controle de utilização ou de buffer.

Para entender melhor, vamos examinar o controle de utilização e o buffer.

Controle de utilização: se a origem da demanda tiver capacidade de repetição, você poderá implementar o controle de utilização. O controle de utilização informa à origem que, se não for possível atender à solicitação no momento, ela deverá tentar novamente mais tarde. A origem espera por um período e repete a solicitação. A implementação do controle de utilização tem a vantagem de limitar a quantidade máxima de recursos e custos da workload. Na AWS, o [Amazon API Gateway](#) pode ser usado para implementar o controle de utilização.

Buffer: uma abordagem baseada em buffer usa produtores (componentes que enviam mensagens para a fila), consumidores (componentes que recebem mensagens da fila) e uma fila (que contém mensagens) para armazenar as mensagens. As mensagens são lidas pelos consumidores e processadas, permitindo que as mensagens sejam executadas na taxa que atenda aos requisitos de negócios dos consumidores. Usando uma metodologia centrada em buffer, as mensagens dos produtores são armazenadas em filas ou fluxos, prontas para serem acessadas pelos consumidores em um ritmo alinhado às demandas operacionais.

Na AWS, você pode escolher entre vários serviços para implementar uma abordagem de buffering. O [Amazon Simple Queue Service \(Amazon SQS\)](#) é um serviço gerenciado que fornece filas que permitem que um único consumidor leia mensagens individuais. O [Amazon Kinesis](#) oferece um fluxo que permite que muitos consumidores leiam as mesmas mensagens.

O buffer e o controle de utilização podem suavizar qualquer pico modificando a demanda da workload. Use o controle de utilização quando os clientes repetirem ações e use o buffer para reter a solicitação e processá-la posteriormente. Ao trabalhar com uma arquitetura com uma abordagem baseada em buffer, arquitecte a workload para atender à solicitação no tempo necessário e verifique se é possível lidar com solicitações duplicadas de trabalho. Analise a demanda geral, a taxa de alteração e o tempo de resposta necessário para dimensionar adequadamente o controle ou buffer necessário.

Etapas de implementação

- Analise os requisitos do cliente: analise as solicitações do cliente para determinar se são capazes de executar novas tentativas. Para clientes que não podem realizar novas tentativas, será necessário implementar buffers. Analise a demanda geral, a taxa de alteração e o tempo de resposta necessário para determinar o tamanho do controle de utilização ou do buffer necessário.
- Implemente um buffer ou controle de utilização: implemente um buffer ou controle de utilização na workload. Uma fila como o Amazon Simple Queue Service (Amazon SQS) pode fornecer um buffer para os componentes da workload. O Amazon API Gateway pode oferecer controle de utilização para componentes da workload.

Recursos

Práticas recomendadas relacionadas:

- [SUS02-BP06 Implementar armazenamento em buffer ou controle de utilização para nivelar a curva da demanda](#)
- [REL05-BP02 Controlar a utilização de solicitações](#)

Documentos relacionados:

- [AWS Auto Scaling](#)
- [Agendador de instâncias da AWS](#)
- [Amazon API Gateway](#)

- [Amazon Simple Queue Service](#)
- [Conceitos básicos do Amazon SQS](#)
- [Amazon Kinesis](#)

Vídeos relacionados:

- [Escolher o serviço de mensagens certo para a aplicação distribuída](#)

Exemplos relacionados:

- [Gerenciar e monitorar o controle de utilização de APIs em suas workloads](#)
- [Controlar a utilização de uma API REST em camadas e multilocatário em grande escala usando o API Gateway](#)
- [Habilitar a hierarquização e o controle de utilização em uma solução SaaS multilocatária do Amazon EKS usando o Amazon API Gateway](#)
- [Integrar aplicações usando filas e mensagens](#)

COST09-BP03 Fornecer recursos dinamicamente

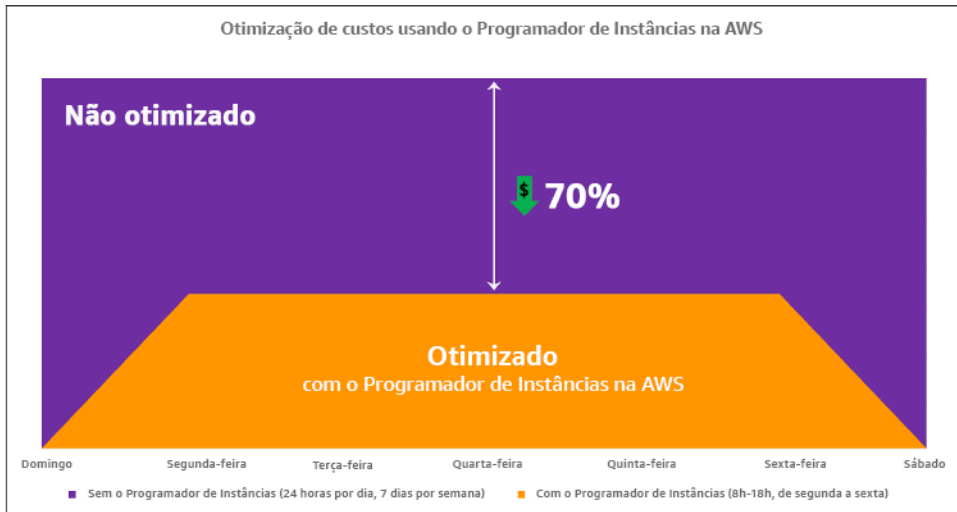
Os recursos são provisionados de maneira planejada. Isso pode ser feito com base na demanda, como por meio do ajuste de escala automático, ou com base no tempo, em que a demanda é previsível e os recursos são fornecidos em função do tempo. Esses métodos ocasionam a menor quantidade de superprovisionamento ou subprovisionamento.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Há várias maneiras como os clientes da AWS podem aumentar os recursos disponíveis para suas aplicações e fornecer recursos para atender à demanda. Uma dessas opções é usar o AWS Instance Scheduler, que automatiza o início e a interrupção de instâncias do Amazon Elastic Compute Cloud (Amazon EC2) e do Amazon Relational Database Service (Amazon RDS). A outra opção é usar o AWS Auto Scaling, que possibilita escalar automaticamente seus recursos de computação com base na demanda da aplicação ou do serviço. O fornecimento de recursos com base na demanda permitirá que você pague somente pelos recursos utilizados, reduza os custos lançando recursos quando eles forem necessários e os encerre quando não forem.

O [Agendador de instâncias da AWS](#) permite configurar a interrupção e o início de suas instâncias do Amazon EC2 e do Amazon RDS em horários definidos para que você possa atender à demanda pelos mesmos recursos em um padrão de tempo consistente, por exemplo, acesso diário dos usuários às instâncias do Amazon EC2 às 8h, que não são necessárias após as 18h. Essa solução ajuda a reduzir o custo operacional interrompendo recursos que não estão sendo usados e iniciá-los quando eles são necessários.



Otimização de custos com o Agendador de instâncias da AWS.

Você também pode configurar facilmente programações para suas instâncias do Amazon EC2 em suas contas e regiões com uma interface de usuário (IU) simples usando a Configuração rápida do AWS Systems Manager. É possível programar instâncias do Amazon EC2 e do Amazon RDS com o Agendador de instâncias da AWS e interromper e iniciar instâncias existentes. No entanto, não é possível parar e iniciar instâncias que fazem parte do seu grupo do Auto Scaling (ASG) ou que gerenciam serviços como o Amazon Redshift ou o Amazon OpenSearch Service. Os grupos do Auto Scaling têm seu próprio agendamento para as instâncias do grupo e essas instâncias são criadas.

O [AWS Auto Scaling](#) ajuda você a ajustar sua capacidade para manter uma performance estável e previsível pelo menor custo possível para atender às variações de demanda. Trata-se de um serviço totalmente gerenciado e gratuito para escalar a capacidade da aplicação e que se integra às instâncias do Amazon EC2 e às frotas spot, ao Amazon ECS, ao Amazon DynamoDB e ao Amazon Aurora. O Auto Scaling oferece descoberta automática de recursos para ajudar a encontrar recursos na sua workload que possam ser configurados, tem estratégias de ajuste de escala incorporadas para otimizar performance, custos ou um equilíbrio entre os dois, além de oferecer ajuste de escala preditivo para ajudar com picos que ocorrem regularmente.

Há várias opções de ajuste de escala disponíveis para escalar seu grupo do Auto Scaling:

- Manter níveis de instâncias atuais em todos os momentos
- Dimensionar manualmente
- Escala baseada em uma programação
- Escala com base em demanda
- Usar o ajuste de escala preditivo

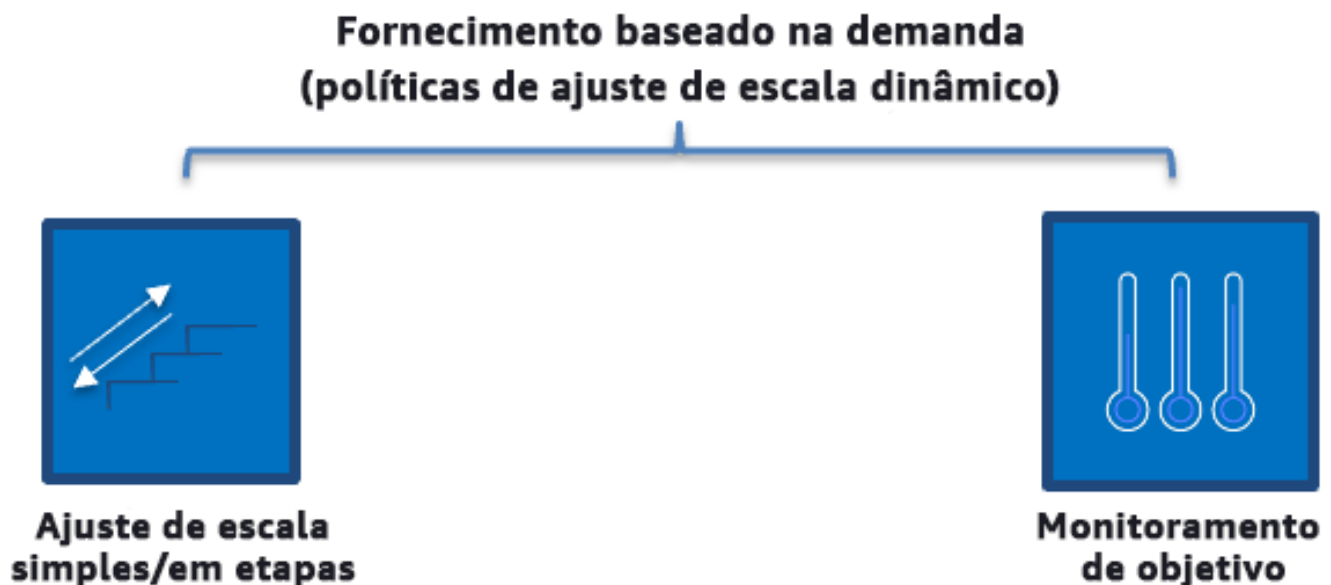
As políticas do Auto Scaling diferem e podem ser categorizadas como políticas de ajuste de escala dinâmicas e agendadas. As políticas dinâmicas são ajuste de escala manual ou dinâmico, ajuste de escala agendado ou preditivo. Você pode usar políticas para ajuste de escala dinâmico, agendado e preditivo. Você também pode usar métricas e alarmes do [Amazon CloudWatch](#) para acionar eventos de ajuste de escala para sua workload. Recomendamos usar [modelos de execução](#) para garantir que esteja acessando os recursos e melhorias mais recentes. Nem todos os recursos do Auto Scaling estão disponíveis quando você usa configurações de execução. Por exemplo, não é possível criar um grupo do Auto Scaling que execute instâncias spot e sob demanda ou que especifique vários tipos de instância. Você deve usar um modelo de execução para configurar esses recursos. Ao usar modelos de execução, recomendamos versionar cada um. Com o versionamento dos modelos de execução, você pode criar um subconjunto do conjunto completo de parâmetros. Em seguida, você pode reutilizá-lo para criar outras versões do mesmo modelo de execução.

É possível usar o AWS Auto Scaling ou incorporar ajuste de escala em seu código com as [AWS APIs ou SDKs](#). Isso reduz os custos gerais da workload removendo o custo operacional de fazer alterações manualmente em seu ambiente, e quaisquer alterações podem ser realizadas muito mais rapidamente. Isso também atende à mobilização de recursos da workload de acordo com sua demanda a qualquer momento. Para seguir essa prática recomendada e fornecer recursos de forma dinâmica para sua organização, você precisa entender os ajustes de escala horizontal e vertical na Nuvem AWS, bem como a natureza das aplicações executadas em instâncias do Amazon EC2. É melhor para sua equipe de gerenciamento financeiro na nuvem trabalhar com equipes técnicas a fim de seguir essa prática recomendada.

O [Elastic Load Balancing \(ELB\)](#) ajuda você a escalar distribuindo a demanda entre vários recursos. Com o uso do ASG e do Elastic Load Balancing, você pode gerenciar as solicitações recebidas roteando o tráfego de forma ideal para que nenhuma instância fique sobrecarregada em um grupo do Auto Scaling. As solicitações seriam distribuídas entre todos os destinos de um grupo-alvo de forma contínua, sem considerar a capacidade nem a utilização.

As métricas típicas podem ser métricas padrão do Amazon EC2, como utilização de CPU, throughput de rede e latência de solicitação/resposta observada pelo Elastic Load Balancing. Quando possível, use uma métrica que seja indicativa da experiência do cliente. Normalmente é uma métrica personalizada que pode se originar do código da aplicação em sua workload. Para elaborar como atender à demanda dinamicamente neste documento, vamos agrupar o Auto Scaling em duas categorias, como modelos de fornecimento baseados na demanda e baseados no tempo, e nos aprofundarmos em cada uma delas.

Fornecimento baseado em demanda: utilize a elasticidade da nuvem para fornecer recursos para atender às mudanças na demanda e depender do estado de demanda quase em tempo real. Para fornecimento baseado em demanda, use as APIs ou os recursos de serviço para variar programaticamente a quantidade de recursos de nuvem em sua arquitetura. Isso permite que você ajuste a escala de componentes em sua arquitetura e aumente o número de recursos durante picos de demanda a fim de manter a performance e reduzir a capacidade quando a demanda diminui para reduzir os custos.

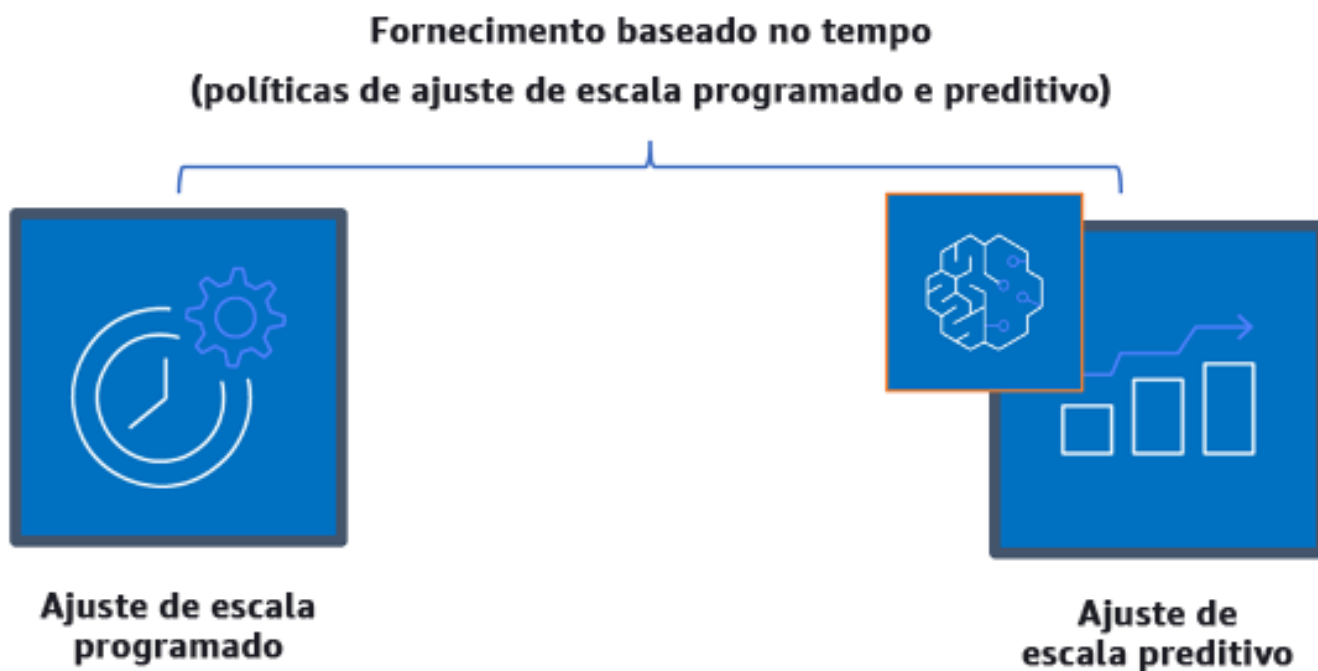


Políticas de ajuste de escala dinâmico com base na demanda

- Ajuste de escala simples/em etapas: monitora métricas e adiciona/remove instâncias de acordo com as etapas definidas manualmente pelos clientes.
- Monitoramento do objetivo: mecanismo de controle semelhante a um termostato que adiciona ou remove instâncias automaticamente para manter as métricas em um objetivo definido pelo cliente.

Ao arquitetar com uma abordagem baseada em demanda, tenha em mente dois pontos essenciais. Primeiro, entenda a rapidez com que você deve provisionar novos recursos. Segundo, entenda que o tamanho da margem entre fornecimento e demanda mudará. Você deve estar pronto para lidar com a taxa de alteração na demanda e também estar pronto para falhas de recursos.

Fornecimento com base em tempo: uma abordagem baseada em tempo alinha a capacidade de recursos à demanda previsível ou bem definida em relação ao tempo. Essa abordagem costuma não depender dos níveis de utilização dos recursos. Uma abordagem baseada em tempo garante que os recursos estejam disponíveis no momento específico em que são necessários e podem ser fornecidos sem nenhum atraso devido a procedimentos de inicialização e verificações do sistema ou de consistência. Usando uma abordagem baseada em tempo, você pode fornecer recursos adicionais ou aumentar a capacidade durante períodos ocupados.



Políticas de ajuste de escala baseado em tempo

Você pode usar o ajuste de escala automático agendado ou preditivo para implementar uma abordagem baseada em tempo. As workloads podem ser agendadas para aumentar ou reduzir a escala horizontalmente em horários definidos (por exemplo, o início do horário comercial), tornando os recursos disponíveis quando os usuários chegarem ou a demanda aumentar. O ajuste de escala preditivo usa padrões para aumentar a escala horizontalmente enquanto o ajuste de escala agendado usa horários predefinidos para isso. Você também pode usar uma [estratégia de](#)

[seleção de tipo de instância baseada em atributo \(ABS\)](#) em grupos do Auto Scaling, o que permite expressar seus requisitos de instância como um conjunto de atributos, como vCPU, memória e armazenamento. Isso também permite usar automaticamente os tipos de instância de geração mais recente quando eles são lançados e acessar uma variedade mais ampla de capacidade com instâncias spot do Amazon EC2. O Amazon EC2 Fleet e o Amazon EC2 Auto Scaling selecionam e executam instâncias que se ajustam aos atributos especificados, eliminando a necessidade de escolher manualmente os tipos de instância.

Você também pode aproveitar as [APIs e os SDKs da AWS](#) e o [AWS CloudFormation](#) para provisionar e desativar automaticamente ambientes inteiros conforme necessário. Essa abordagem é adequada para ambientes de desenvolvimento ou teste que são executados apenas em períodos ou horários comerciais definidos. Você pode usar APIs para ajustar a escala dos recursos dentro de um ambiente (ajuste de escala vertical). Por exemplo, você pode escalar uma workload de produção alterando o tamanho ou a classe da instância. Isso pode ser feito interrompendo e iniciando a instância e selecionando a classe ou o tamanho da instância diferente. Essa técnica também pode ser aplicada a outros recursos, como Volumes Elásticos do Amazon EBS, que podem ser modificados para aumentar o tamanho, ajustar a performance (IOPS) ou alterar o tipo de volume durante o uso.

Ao arquitetar com uma abordagem baseada em tempo, tenha em mente dois pontos essenciais. Primeiro, qual é a consistência do padrão de uso? Segundo, qual será o impacto se o padrão mudar? Você pode aumentar a precisão das previsões monitorando suas workloads e usando inteligência de negócios. Se você vir alterações significativas no padrão de uso, poderá ajustar os tempos para garantir que a cobertura seja fornecida.

Etapas de implementação

- Configure o ajuste de escala agendado: para alterações previsíveis na demanda, o ajuste de escala com base em tempo pode fornecer a quantidade correta de recursos em tempo hábil. Ele também será útil se a criação e a configuração de recursos não forem rápidas o suficiente para responder a alterações na demanda. Usando a análise de workload, configure o ajuste de escala agendado usando o AWS Auto Scaling. Para configurar o cronograma baseado em tempo, você pode usar o ajuste de escala preditivo ou agendado para aumentar o número de instâncias do Amazon EC2 em seus grupos do Auto Scaling com antecedência de acordo com as alterações de carga esperadas ou previsíveis.
- Configure o ajuste de escala preditivo: o ajuste de escala preditivo permite aumentar o número de instâncias do EC2 em seu grupo do Auto Scaling em antecipação aos padrões diários e semanais nos fluxos de tráfego. Se você tem picos de tráfego regulares e aplicações que demoram muito

para ser iniciadas, considere usar o ajuste de escala preditivo. O ajuste de escala preditivo pode ajudar você a escalar com maior rapidez inicializando a capacidade antes da carga projetada em comparação com o ajuste de escala dinâmico isolado, que é reativo por natureza. Por exemplo, se os usuários começarem a usar sua workload no início do horário comercial e não usá-la após o expediente, o ajuste de escala preditivo poderá adicionar capacidade antes do horário comercial, o que elimina o atraso do ajuste de escala dinâmico para reagir a mudanças no tráfego.

- Configure o ajuste de escala automático dinâmico: para configurar o ajuste de escala com base em métricas de workload ativas, use o Auto Scaling. Use a análise e configure o Auto Scaling para iniciar nos níveis de recursos corretos e garanta que a workload escale no tempo necessário. Você pode iniciar e escalar automaticamente uma frota de instâncias sob demanda e instâncias spot em um único grupo do Auto Scaling. Além de receber descontos pelo uso de instâncias spot, você pode usar instâncias reservadas ou um Savings Plan para receber taxas de desconto da definição de preço normal de instância sob demanda. Todos esses fatores combinados ajudam você a otimizar sua redução de custos para instâncias do Amazon EC2 e a obter a escala e a performance desejadas para a aplicação.

Recursos

Documentos relacionados:

- [AWS Auto Scaling](#)
- [Agendador de instâncias da AWS](#)
- Escalar o tamanho do grupo do Auto Scaling
- [Conceitos básicos do Amazon EC2 Auto Scaling](#)
- [Conceitos básicos do Amazon SQS](#)
- [Ajuste de escala agendado para o Amazon EC2 Auto Scaling](#)
- [Escala preditiva para o Amazon EC2 Auto Scaling](#)

Vídeos relacionados:

- [Políticas de ajuste de escala com monitoramento do objetivo para Auto Scaling](#)
- [Agendador de instâncias da AWS](#)

Exemplos relacionados:

- [Seleção de tipo de instância baseada em atributos para Auto Scaling para Amazon EC2 Fleet](#)
- [Otimizar o Amazon Elastic Container Service para custos usando ajuste de escala agendado](#)
- [Ajuste de escala preditivo com o Amazon EC2 Auto Scaling](#)
- [Como faço para usar o Agendador de instâncias com o AWS CloudFormation para agendar as instâncias do Amazon EC2?\)](#)

Otimização ao longo do tempo

Na AWS, é possível fazer otimizações ao longo do tempo revisando e implementando novos serviços em sua workload.

À medida que a AWS lança novos serviços e recursos, é prática recomendada revisar suas decisões de arquitetura atuais para garantir que elas continuem oferecendo um bom custo-benefício. Conforme seus requisitos mudam, seja agressivo na desativação de recursos, componentes e workloads de que não precisa mais. Considere as práticas recomendadas a seguir para ajudar você a fazer otimizações ao longo do tempo.

Ao otimizar suas workloads ao longo do tempo e melhorar a cultura de [CFM](#) em sua organização, avalie o custo do esforço das operações na nuvem, revise suas operações demoradas na nuvem e automatize-as para reduzir os esforços humanos e os custos adotando serviços da AWS relacionados, produtos de terceiros ou ferramentas personalizadas (como [AWS CLI](#) ou [AWS SDKs](#)).

Tópicos

- [Definir um processo de revisão e analisar sua workload com frequência](#)
- [Automatização de operações](#)

Definir um processo de revisão e analisar sua workload com frequência

Práticas recomendadas

- [COST10-BP01 Desenvolver um processo de revisão de workloads](#)
- [COST10-BP02 Revisar e analisar a workload regularmente](#)

COST10-BP01 Desenvolver um processo de revisão de workloads

Desenvolva um processo que defina os critérios e o processo para a revisão de workloads. O esforço de análise deve refletir o benefício potencial. Por exemplo, workloads principais ou workloads com valor superior a 10% da fatura são revisadas trimestralmente ou a cada seis meses, enquanto workloads abaixo de 10% são revisadas anualmente.

Nível de risco exposto se esta prática recomendada não for estabelecida: Alto

Orientação para implementação

Para manter a workload mais econômica, é necessário revisá-la regularmente para saber se há oportunidades de implementar novos serviços, recursos e componentes. Para obter custos gerais mais baixos, o processo deve ser proporcional à quantidade potencial de economia. Por exemplo, as workloads que representam 50% do seu gasto geral devem ser revisadas com mais frequência e mais precisão do que as workloads que representam 5% do seu gasto geral. Leve em consideração quaisquer fatores externos ou volatilidade. Se a workload atender a uma área geográfica ou segmento de mercado específico e houver previsão de mudanças nessa área, revisões mais frequentes poderão resultar em economias de custos. Outro fator em questão é o esforço para implementar alterações. Se houver custos significativos em testes e validação de alterações, as revisões deverão ser menos frequentes.

Leve em consideração o custo de longo prazo de manutenção de componentes e recursos obsoletos e a incapacidade de implementar novos recursos neles. O custo atual de testes e validação pode exceder o benefício proposto. No entanto, ao longo do tempo, o custo de fazer a mudança pode aumentar significativamente à medida que a lacuna entre a workload e as tecnologias atuais aumenta, resultando em custos ainda maiores. Por exemplo, o custo da migração para uma nova linguagem de programação pode não ser econômico no momento. No entanto, em cinco anos, o custo de pessoas com qualificações nessa linguagem pode aumentar e, devido ao crescimento da workload, você estaria movendo um sistema ainda maior para a nova linguagem, exigindo ainda mais esforço do que anteriormente.

Divida sua workload em componentes, atribua o custo do componente (uma estimativa é suficiente) e liste os fatores (por exemplo, esforço e mercados externos) ao lado de cada componente. Use esses indicadores para determinar uma frequência de revisão para cada workload. Por exemplo, você pode ter servidores web como um alto custo, baixo esforço de alteração e fatores externos elevados, o que resulta em uma alta frequência de revisão. Um banco de dados central pode ser de custo médio, alto esforço de alteração e baixos fatores externos, resultando em uma média frequência de análise.

Defina um processo para avaliar novos serviços, padrões de design, tipos de recursos e configurações para otimizar o custo de sua workload conforme ficarem disponíveis. Semelhante aos processos de [revisão do pilar Performance](#) e [revisão do pilar Confiabilidade](#), identifique, valide e priorize as atividades de otimização e melhoria e a remediação de problemas e incorpore-as à sua lista de pendências.

Etapas de implementação

- Defina a frequência de revisões: defina a frequência com que a workload e os respectivos componentes devem ser revisados. Aloque tempo e recursos para o aprimoramento contínuo e analise a frequência para melhorar a eficiência e a otimização da sua workload. Essa é uma combinação de fatores e pode diferir de workload para workload em sua organização e entre componentes na workload. Os fatores comuns incluem: a importância para a organização medida em termos de receita ou marca, o custo total da execução da workload (incluindo custos operacionais e de recursos), a complexidade da workload, a facilidade da implementação de uma alteração, qualquer contrato de licenciamento de software e se uma alteração geraria aumentos significativos nos custos de licenciamento devido a licenciamento punitivo. Os componentes podem ser definidos de maneira funcional ou técnica, como bancos de dados e servidores Web ou recursos de computação e armazenamento. Equilibre os fatores de acordo e desenvolva um período para a workload e os respectivos componentes. Você pode decidir analisar a workload completa a cada 18 meses, analisar os servidores Web a cada seis meses, o banco de dados a cada doze meses, a computação e o armazenamento de curto prazo a cada seis meses e o armazenamento de longo prazo a cada doze meses.
- Defina o rigor da revisão: defina quanto esforço é gasto na revisão da workload ou dos respectivos componentes. Semelhante à frequência da análise, esse é um equilíbrio de vários fatores. Avalie e priorize oportunidades de melhorias para concentrar os esforços nos locais onde eles oferecem os maiores benefícios enquanto calcula quanto esforço é necessário para essas atividades. Se os resultados esperados não satisfizerem as metas e o esforço necessário custar mais, itere usando cursos de ação alternativos. Seus processos de análise devem incluir tempo e recursos dedicados para possibilitar melhorias incrementais contínuas. Por exemplo, você pode decidir gastar uma semana de análise no componente do banco de dados, uma semana de análise para recursos computacionais e quatro horas para análises de armazenamento.

Recursos

Documentos relacionados:

- [Notícias do blog da AWS](#)
- [Tipos de computação em nuvem](#)
- [Novidades da AWS](#)

Exemplos relacionados:

- [Serviços proativos do AWS Support](#)

- [Revisões regulares da workload para workloads do SAP](#)

COST10-BP02 Revisar e analisar a workload regularmente

As workloads existentes são revisadas regularmente com base em cada processo definido para descobrir se é possível adotar novos serviços, substituir serviços já em vigor ou refazer a arquitetura das workloads.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A AWS adiciona constantemente novos recursos para que você possa experimentar e novar mais rápido com a tecnologia mais recente. [AWS Novidades](#) detalha como a AWS está fazendo isso e fornece uma visão geral rápida dos serviços, recursos e anúncios de expansão regional da AWS à medida que são lançados. Você pode examinar detalhadamente os lançamentos que foram anunciados e usá-los para revisar e analisar suas workloads existentes. Para obter os benefícios de novos serviços e recursos da AWS, analise suas workloads e implemente novos serviços e recursos conforme necessário. Isso significa que você pode precisar substituir os serviços que você usa para a workload ou modernizar a workload para adotar novos serviços da AWS. Por exemplo, você pode revisar suas workloads e substituir o componente de mensagens pelo Amazon Simple Email Service. Fazer isso elimina os custos de operação e manutenção de uma frota de instâncias e, ao mesmo tempo, fornece toda a funcionalidade a um custo reduzido.

Para analisar sua workload e destacar possíveis oportunidades, considere não apenas novos serviços, mas também novas formas de criar soluções. Veja os vídeos [Esta é a minha arquitetura](#) na AWS para saber mais sobre os projetos de arquitetura de outros clientes, seus desafios e suas soluções. Confira a [série All-In](#) para descobrir aplicações reais de serviços da AWS e histórias de clientes. Você também pode assistir à série de vídeos [De volta ao básico](#) que explica, examina e detalha as práticas recomendadas básicas para padrões de arquitetura de nuvem. Outra fonte são os vídeos [Como fazer isto](#), criados para ajudar pessoas com grandes ideias sobre como dar vida a seu produto mínimo viável (MVP) usando serviços da AWS. Desse modo, criadores do mundo inteiro que tiverem uma grande ideia poderão obter orientações arquiteturais de arquitetos de soluções da AWS experientes. Finalmente, você pode revisar os materiais de recursos de [Conceitos básicos](#), os quais oferecem tutoriais passo a passo.

Antes de iniciar seu processo de avaliação, siga os requisitos de sua empresa com relação a workload, segurança e privacidade dos dados para usar requisitos específicos de serviço ou de região e performance e, ao mesmo tempo, siga o processo de avaliação que foi acordado.

Etapas de implementação

- Revise regularmente a workload: usando o processo definido, execute análises com a frequência especificada. Verifique se você despendeu a quantidade correta de esforço em cada componente. Esse processo seria semelhante ao processo de design inicial em que você selecionou serviços para otimização de custos. Analise os serviços e os benefícios que eles trariam, esse fator de tempo no custo de fazer a mudança, e não apenas os benefícios de longo prazo.
- Implemente novos serviços: se o resultado da análise for implementar alterações, primeiro execute uma linha de base da workload para saber o custo atual por saída. Implemente as alterações e, em seguida, execute uma análise para confirmar o novo custo por saída.

Recursos

Documentos relacionados:

- [Notícias do blog da AWS](#)
- [Novidades da AWS](#)
- [Documentação do AWS](#)
- [Conceitos básicos da AWS](#)
- [Recursos gerais da AWS](#)

Vídeos relacionados:

- [AWS: Esta é a minha arquitetura](#)
- [AWS: De volta ao básico](#)
- [AWS: Série All-In](#)
- [Como fazer isto](#)

Automatização de operações

Práticas recomendadas

- [COST11-BP01 Realizar automações nas operações](#)

COST11-BP01 Realizar automações nas operações

Avalie os custos operacionais na nuvem, concentrando-se em quantificar a economia de tempo e de esforço em tarefas administrativas e implantações, bem como em mitigar o risco de erros humanos, conformidade e outras operações por meio da automação. Avalie os custos associados e o tempo necessários para os esforços operacionais e implemente a automação de tarefas administrativas para minimizar o esforço manual sempre que possível.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

A automação de operações reduz a frequência das tarefas manuais, melhora a eficiência e beneficia os clientes por oferecer uma experiência consistente e confiável em implantação, administração ou operação de workloads. Você pode liberar recursos de infraestrutura de tarefas operacionais manuais e usá-los para tarefas e inovações de maior valor, o que melhora os resultados dos negócios. As empresas necessitam de um método comprovado e testado para gerenciar suas workloads na nuvem. Essa solução deve ser segura, rápida e econômica, com risco mínimo e e máxima confiabilidade.

Primeiro, priorize suas atividades operacionais com base no esforço necessário examinando o custo geral das operações. Por exemplo, quanto tempo é preciso para implantar novos recursos na nuvem, realizar alterações de otimização nos recursos existentes ou implementar as configurações necessárias? Examine o custo total das ações humanas, incluindo o custo de operações e gerenciamento como fator. Priorize a automação das tarefas administrativas para reduzir o esforço humano.

A avaliação do esforço deve refletir o provável benefício. Por exemplo, examine o tempo gasto na execução de tarefas manuais em contraposição a tarefas automáticas. Priorize a automatização de atividades repetitivas, de alto valor, demoradas e complexas. As atividades que apresentam um alto valor ou alto risco de erro humano normalmente são o melhor lugar para começar a automatizar, pois, com frequência, o risco cria um custo operacional adicional não desejado (como horas extras de trabalho da equipe de operações).

Use ferramentas de automação, como o AWS Systems Manager ou o AWS Config para racionalizar processos de operações, conformidade, monitoramento, ciclo de vida e encerramento. Com serviços

e ferramentas da AWS, além de produtos de terceiros, você pode personalizar as automações implementadas para atender às suas necessidades específicas. A tabela abaixo mostra alguns recursos e funções essenciais de operações que você pode obter com os serviços da AWS para automatizar a administração e a operação:

- [AWS Audit Manager](#): audite continuamente o uso da AWS para simplificar a avaliação de risco e conformidade.
- [AWS Backup](#): gerencie e automatize centralmente a proteção de dados para o Amazon.
- [AWS Config](#): configure recursos computacionais, avalie, audite e analise configurações e o inventário de recursos.
- [AWS CloudFormation](#): inicie recursos altamente disponíveis com a infraestrutura como código.
- [AWS CloudTrail](#): gerenciamento, conformidade e controle de mudanças de TI.
- Agende eventos do [Amazon EventBridge](#) e acione o AWS Lambda para adotar medidas.
- [AWS Lambda](#): automatize processos repetitivos acionando-os com eventos ou executando-os em um cronograma fixo com o AWS EventBridge.
- [AWS Systems Manager](#): inicie e interrompa workloads, corrija sistemas operacionais, automatize a configuração e gerencie continuamente.
- [AWS Step Functions](#): Agende trabalhos e automatize fluxos de trabalho.
- [AWS Service Catalog](#): Consumo de modelos, infraestrutura como código com conformidade e controle.

Se você quiser adotar automações imediatamente com o uso de produtos e serviços da AWS e não contar com as habilidades necessárias em sua organização, entre em contato com [AWS Managed Services \(AMS\)](#), a [AWS Professional Services](#) ou [parceiros da AWS](#) para aumentar a adoção de automação e melhorar sua excelência operacional na nuvem.

A AWS Managed Services (AMS) é um serviço que opera a infraestrutura da AWS em nome de clientes e parceiros corporativos. Ele fornece um ambiente seguro e compatível no qual você pode implantar as workloads. O AMS usa modelos operacionais de nuvem empresarial com automação para permitir que você atenda aos requisitos da organização, migre para a nuvem mais rapidamente e reduza os custos de gerenciamento constantes.

A AWS Professional Services também pode ajudar você a alcançar os resultados de negócios desejados e a automatizar as operações com a AWS. Essa equipe ajuda os cliente a implantar operações de TI automatizadas, robustas e ágeis, bem como recursos de governança otimizados

para a nuvem. Para ver exemplos detalhados de monitoramento e práticas recomendadas, consulte o whitepaper [Pilar Excelência operacional](#).

Etapas de implementação

- Crie uma vez e implante várias: use infraestrutura como código, como CloudFormation, AWS SDK ou AWS CLI para implantar uma vez e usar várias vezes em ambientes semelhantes ou em cenários de recuperação de desastres. Marque com tags enquanto monitora o consumo conforme definido em outras práticas recomendadas. Use o [AWS Launch Wizard](#) para reduzir o tempo de implantação de muitas workloads corporativas populares. A AWS Launch Wizard orienta você no dimensionamento, na configuração e na implantação de workloads corporativas de acordo com as práticas recomendadas da AWS. Você também pode usar o [Service Catalog](#), que ajuda a criar e gerenciar modelos aprovados de infraestrutura como código para uso na AWS para que qualquer pessoa possa descobrir recursos de nuvem de autoatendimento aprovados.
- Automatize a conformidade contínua: considere automatizar a avaliação e a correção das configurações registradas em relação a padrões predefinidos. Ao combinar o AWS Organizations com os recursos do AWS Config e [AWS CloudFormation](#), você pode gerenciar e automatizar com eficiência a conformidade da configuração em grande escala para centenas de contas-membro. É possível analisar as mudanças nas configurações e nas relações entre os recursos da AWS e examinar em detalhe o histórico da configuração de um recurso.
- Automatização das tarefas de monitoramento: a AWS fornece várias ferramentas que você pode usar para monitorar serviços. É possível configurar essas ferramentas para automatizar as tarefas de monitoramento. Crie e implemente um plano de monitoramento que colete dados de todas as partes da workload para que você possa depurar com maior facilidade uma falha multiponto, caso ocorra. Por exemplo, é possível usar as ferramentas de monitoramento automatizadas para observar o Amazon EC2 e informar quando algo está errado nas verificações de status do sistema, nas verificações de status de instâncias e nos alarmes do Amazon CloudWatch.
- Automatize a manutenção e as operações: execute operações de rotina automaticamente sem intervenção humana. Usando serviços e ferramentas da AWS, você pode escolher quais automações da AWS deverão ser implementadas e personalizadas de acordo com seus requisitos específicos. Por exemplo, use o [EC2 Image Builder](#) para criar, testar e implantar imagens de máquinas virtuais e de contêineres para uso na AWS ou on-premises ou para aplicar patches em suas instâncias do EC2 com o AWS SSM. Se a ação desejada não puder ser realizada com serviços da AWS, ou se você precisar de ações mais complexas com recursos de filtragem, automatize suas operações usando o [AWS Command Line Interface](#) (AWS CLI) ou as ferramentas do AWS SDK. A AWS CLI oferece a capacidade de automatizar todo o processo de controle e gerenciamento de serviços da AWS com scripts sem usar o AWS Management Console. Selecione

seus AWS SDKs preferidos para interagir com os serviços da AWS. Para outros exemplos de código, consulte o [repositório de exemplos](#) de código do AWS SDK.

- Crie um ciclo de vida contínuo com automações: é importante que você estabeleça e preserve políticas maduras de ciclo de vida, não apenas para regulamentações ou redundância, mas também para otimização de custos. É possível usar o AWS Backup para gerenciar e automatizar centralmente a proteção de datastores, como buckets, volumes, bancos de dados e sistemas de arquivos. Você também pode usar o Amazon Data Lifecycle Manager para automatizar a criação, a retenção e a exclusão de snapshots do EBS e de AMIs apoiadas pelo EBS.
- Exclua recursos desnecessários: é muito comum acumular recursos não utilizados no sandbox ou em Contas da AWS de desenvolvimento. Os desenvolvedores criam e experimentam vários serviços e recursos como parte do ciclo normal de desenvolvimento, mas não excluem esses recursos quando não são mais necessários. Recursos não utilizados podem gerar custos desnecessários e, às vezes, altos para a organização. A exclusão desses recursos pode reduzir os custos de operação desses ambientes. Certifique-se de que seus dados não sejam necessários. Se não tiver certeza, faça backup. Você pode usar o AWS CloudFormation para limpar as pilhas implantadas, o que exclui automaticamente a maioria dos recursos definidos no modelo. Como alternativa, você pode criar uma automação para a exclusão de recursos da AWS usando ferramentas como o [aws-nuke](#).

Recursos

Documentos relacionados:

- [Modernizar as operações na Nuvem AWS](#)
- [Usar serviços da AWS para automação](#)
- [Infraestrutura e automação](#)
- [AWS Systems Manager Automation](#)
- [Monitoramento automatizado e manual](#)
- [Automações da AWS para administração e operações do SAP](#)
- [AWS Managed Services](#)
- [AWS Professional Services](#)

Vídeos relacionados:

- [Automatizar a conformidade contínua em grande escala na AWS](#)

- [Demonstração do AWS Backup: backup entre contas e regiões](#)
- [Correção de patches para suas instâncias do Amazon EC2](#)

Exemplos relacionados:

- [Reinventar as operações automatizadas \(Parte I\)](#)
- [Reinventar as operações automatizadas \(Parte II\)](#)
- [Automatizar a exclusão de recursos da AWS usando o aws-nuke](#)
- [Excluir volumes não utilizados do Amazon EBS com o AWS Config e o AWS SSM](#)
- [Automatizar a conformidade contínua em grande escala na AWS](#)
- [Automações de TI com o AWS Lambda](#)

Conclusão

A otimização de custos e o gerenciamento financeiro na nuvem são um esforço contínuo. Você deve trabalhar regularmente com suas equipes de finanças e tecnologia, analisar sua abordagem arquitetônica e atualizar sua seleção de componentes.

A AWS ajuda você a minimizar o custo enquanto cria implantações altamente resilientes, responsivas e adaptáveis. Para otimizar verdadeiramente o custo da sua implantação, aproveite as ferramentas, as técnicas e as práticas recomendadas discutidas neste documento.

Colaboradores

Os colaboradores deste documento incluem:

- Fatih (Ben) Mergen, líder do pilar Otimização de custos, Well-Architected, Amazon Web Services
- Keith Jarrett, líder de desenvolvimento de negócios, Otimização de custos, Amazon Web Services
- Arthur Basbaum, gerente de desenvolvimento de negócios, Amazon Web Services
- Jarman Hauser, arquiteto comercial, Amazon Web Services

Outras fontes de leitura

Para obter informações adicionais, consulte:

- [Framework Well-Architected da AWS](#)
- [Centro de Arquitetura do AWS](#)

Revisões do documento

Para ser notificado sobre atualizações desse whitepaper, inscreva-se no feed RSS.

Alteração	Descrição	Data
Orientação atualizada sobre práticas recomendadas	Várias atualizações de práticas recomendadas. Nova prática recomendada COST06-BP04.	27 de junho de 2024
Orientação atualizada sobre práticas recomendadas	Atualizações secundárias em práticas recomendadas.	6 de dezembro de 2023
Orientação atualizada sobre práticas recomendadas	As práticas recomendadas foram atualizadas com novas orientações em todo o pilar.	3 de outubro de 2023
Orientação atualizada sobre práticas recomendadas	As práticas recomendadas foram atualizadas com novas diretrizes nas seguintes áreas: Governança , Monitoramento de custos e uso , Seleção do melhor modelo de preços e Gerenciamento de recursos de demanda e fornecimento .	13 de julho de 2023
Atualizações para o novo Framework	Atualizações nas práticas recomendadas com recomendações e adição de novas práticas recomendadas. Pergunta COST 11 adicionada com a nova prática recomendada da COST11-BP01.	10 de abril de 2023
Atualização secundária	A orientação que estava faltando foi restaurada na	13 de janeiro de 2023

	seção sobre modelo de preços.	
Whitepaper atualizado	Práticas recomendadas atualizadas com novas orientações para implementação.	15 de dezembro de 2022
Whitepaper atualizado	Práticas recomendadas ampliadas e planos de melhoria adicionados.	20 de outubro de 2022
Atualização secundária	Adição do pilar Sustentabilidade à introdução.	2 de dezembro de 2021
Atualização secundária	Links atualizados.	25 de abril de 2021
Atualização secundária	Links atualizados.	10 de março de 2021
Atualizações para o novo Framework	Atualizado para incorporar CFM, novos serviços e integração com o Well-Architected.	8 de julho de 2020
Whitepaper atualizado	Atualizado para refletir alterações na AWS e incorporar aprendizados de revisões com clientes.	1º de julho de 2018
Whitepaper atualizado	Atualizado para refletir alterações na AWS e incorporar aprendizados de revisões com clientes.	1 de novembro de 2017
Publicação inicial	Publicação do pilar Otimização de custos: AWS Well-Architected Framework.	1º de novembro de 2016

Avisos

Os clientes são responsáveis por fazer a própria avaliação independente das informações contidas neste documento. Este documento: (a) é apenas para fins informativos, (b) representa as ofertas e práticas de produtos atuais da AWS, que estão sujeitas a alterações sem aviso prévio e (c) não criam nenhum compromisso ou garantia da AWS e de suas afiliadas, fornecedores ou licenciadores. Os produtos ou serviços da AWS são fornecidos “no estado em que se encontram”, sem garantias, representações ou condições de qualquer tipo, expressas ou implícitas. As responsabilidades e as obrigações da AWS com os seus clientes são controladas por contratos da AWS, e este documento não é parte de, nem modifica, qualquer contrato entre a AWS e seus clientes.

© 2023 Amazon Web Services, Inc. ou suas afiliadas. Todos os direitos reservados.

Glossário da AWS

Para obter a terminologia mais recente da AWS, consulte o [glossário da AWS](#) na Referência do Glossário da AWS.